Heterogeneous Effects of Education on Health

Silvia H. Barcellos, University of Southern California

Leandro Carvalho, University of Southern California

Patrick Turley, Harvard University

PRELIMINARY, PLEASE DO NOT CITE WITHOUT PERMISSION

Abstract: In 1972, the mandatory minimum age at which a student could drop out of school in England and Wales was raised from 15 to 16, constraining roughly 15 percent of the student population. We exploit this discontinuous increase in educational attainment to estimate the impact of education on body mass index (BMI) and diabetes approximately 40 years later. While previous literature found no significant effect of education on health, they were not able to investigate whether these effects vary along the distribution of health outcomes. We are able to detect large effects on BMI in the upper quantiles of observed BMI, as large as 2 BMI points at the 90th percentile of BMI, from a baseline of 35.6. Using a genetic predictor of BMI, we also find that those with higher genetic risk of obesity see smaller reductions in BMI as a result of the increase in compulsory schooling while large reductions are seen in those with low genetic risk. Taken together our results point to the importance of considering heterogeneity when estimating the impacts of education on health.

Funding from NIA grant 3P30AG024962-12S1, the USC Population Research Center and NIA grant K01AG050811-01 (Barcellos) is gratefully acknowledged.

I. Introduction

Public policy evaluation research in economics commonly relies on the estimation of mean parameters within a certain population of interest. This reliance on mean effects is usually explained by data or empirical methods limitations or by the lack of credible measures of dimensions in which heterogeneity is believed to exist. While such mean estimates are important and informative to the policy debate, they might offer misguided lessons if the effects of a given policy vary widely across people from different walks of life. Previous research, for example, has shown that mean impacts can miss important distributional effects of welfare reform (Bitler, Gelbach and Hoynes 2006). In this paper we offer evidence that when evaluating the health effects of educational policies it is important to consider (at least) two dimensions of heterogeneity: health and genetic background.

In order to test for heterogeneous effects of education on health we use data from the UK Biobank, an unprecedentedly rich dataset with 500,000+ participants born between 1934 and 1971. Importantly, the data contain continuous physical measures of health and genotypic data, which allows us to document heterogeneity across the distribution of health and genetic background. In this paper, we focus on BMI and genetic predisposition to high BMI. We show that the effects of education significantly vary across the BMI distribution and for people with different genetic predisposition to high BMI, providing evidence that these underexplored sources of heterogeneity are important.

We exploit a natural experiment to overcome the endogeneity of schooling decisions and uncover the causal effects of education on health. England's Raising of School Leaving Age (ROSLA) Order of 1972 increased the minimum school-leaving age from 15 to 16 years. Students born on or after September 1st 1957 had to stay in school until age 16, while students born before that date could drop out at age 15. Using a regression discontinuity design, we estimate that the reform increased average education by 0.15 years. By comparing the health (measured in 2006-2010) of individuals born before and after 9/1/1957, we can determine

whether an increase in secondary schooling had a causal effect on health approximately 40 years later.

Education and health have been shown to be strongly associated in many periods and countries and for a wide range of health measures. Between 2001 and 2014, the richest 1% of men in the US could expect to live 14.6 years more than the poorest 1% (Chetty et al. 2016). Those who are more educated are also healthier while they are alive; they report being in better health and having fewer health conditions and limitations (Cutler and Lleras-Muney 2010 a, b). However, despite the robust empirical association between education and health the role of education as a determinant of health is still debated.

A recent and growing literature exploits changes in compulsory schooling laws to study the causal effects of education on health (e.g. Lleras-Muney 2005; Oreopoulous 2007; Albouy and Lequien 2009; Silles 2009; Powdthavee 2010; Kemptner et al. 2011; Clark and Royer 2013; Jurges et al. 2013). This literature has produced mixed evidence across countries, time, outcomes, and subpopulations, resulting in no consensus on whether the reported correlation reflects causality from more education to better health. Reviewing recent research, Grossman (2015) concludes that "there is enough conflicting evidence (in these recent studies) to warrant more research on the question of whether more schooling does in fact cause better health outcomes."

One critical point that the previous literature failed to address is whether education has heterogeneous effects on health, affecting individuals from different health, social, and genetic backgrounds differently. In other words, the mechanisms that underlie the education-health relationship may be more relevant in some settings and for some populations than others. In particular, some mechanisms may have a positive effect on health while others may have a negative effect. Since the causal "reduced-form" effect of education on health is the result of positive and negative effects, the total effect may be positive, zero or negative depending on which mechanisms are operating in particular populations.

There are many such mechanisms through which education can affect health. First, education affects income and available material resources. Grenet (2013) found that the 1972 ROSLA increased earnings by 6.9% in males who would have dropped out at age 15 otherwise.[1] More material resources, in turn, can mean access to more/better quality health care and a healthier diet, but also more consumption of "bads" such as cigarettes that harm health. Second, education can increase knowledge; the more educated might be better informed about health risks and make better health decisions, including adopting better health behaviors. Goldman and Smith (2002) for example, show compelling evidence that the more educated are more likely to adhere to complex treatment regimens for diabetes and HIV. Cutler and Lleras-Muney (2012) find that the more educated were faster in understanding new evidence of the negative health effects of smoking in the 1950s and 1960s, and in responding to such information by changing smoking behavior. It has also been hypothesized that, by improving status in society, education might reduce stress related to low social rank (see results from the Whitehall studies in Marmot et al. 1978, 1991). Finally, education might affect individual preferences, such as forward looking and risk taking behavior (Becker and Mulligan 1994; Perez-Arce 2011, Barsky et al. 1995).

There are many ways that these relationships described above could be heterogeneous. For example, Carneiro, Heckman, and Vytlacil (2010) find that the financial returns to college attendance can be highly heterogeneous given unobservable characteristics. So if the primary causal pathway from education to health is through material resources, we would expect to see heterogeneous health returns to education. If there is variation in health production functions or health preferences, this heterogeneity will be further augmented. Many factors, including genes, may drive this this heterogeneous relationship. For instance, De Walque (2007) finds that college attendance decreases the chance that an individual will start smoking and increases the

[1] Oreopoulos (2006) estimates the effect of the 1942 ROSLA in the UK, which raised the minimum school leaving age from 14 to 15. He estimates that the returns to school are between 11% and 18%, depending on his specification.

probability that individuals will quit conditional on having started smoking. Bierut et al. (2008) find, however, that genetic markers affect the way the body reacts to nicotine, increasing the number of cigarettes smoked and reducing the likelihood of quitting. Together, this suggests that the interaction of genes and education may have different impacts on health behaviors like smoking.

The remaining sections of this paper are outlined as follows. Section 2 presents the institutional background of the educational system in the UK in 1972 and details of the ROSLA Order. Section 3 describes the data from the UK Biobank used in this analysis, including an explanation of the genetic variables constructed. Section 4 explains the empirical strategy and presents the results for the mean effects of education on health. In section 5, we describe the heterogeneity analysis and its results. Section 6 outlines the strategy to test how the effect of education on health varies by genotype. A discussion of all of these results is found in Section 7. Section 8 concludes.

II. Institutional background

A. The 1972 Raising of School Leaving Age Order

England and Wales' Raising of School Leaving Age (ROSLA) Order of 1972 increased the minimum school-leaving age from 15 to 16 years. A more complete description of the details and historical context of this policy can be found in Clark and Royer (2013), though a brief summary is found here. This reform was unique for using date of birth to determine which students would be affect by the policy. More precisely, students born on or after September 1, 1957 were required to remain in school until they turned 16 while students born before that date could drop out at age 15. In our data, this change forced roughly 15% of students to get an additional year of education.

The 1972 ROSLA was not unanticipated. The 1944 Education Act, which raised the minimum school-leaving age to 15 years, included a provisions that the minimum age could be raised to 16 years in the future when the Minister of Education was convinced that such a

change was possible to implement. In 1964, preparations began to be made, primarily involving building up the infrastructure to absorb the additional students. After 8 years, the ROSLA went into effect.

B. Qualifications

In addition to directly increasing students' years of education, the ROSLA may have also given the students the chance to obtain qualifications they may not have received otherwise. The General Certificate of Education (GCE) Ordinary Level, also called the O-level or O level, was a school-leaving qualification offered in the UK between 1951 and 1987. The O-level was the typical examination taken by sixteen-year olds in grammar schools, which are more academically oriented. The O-level was predominantly exam-based with each subject as a separate O-level in its own right. The grading of the O-level changed over time but an O-level at grades A-C corresponded to a Level 2 qualification.

The Certificate of Secondary Education (CSE) was a school-leaving qualification offered in the UK between 1965 and 1987. CSE courses and examinations were designed for students who were not thought likely to pass the O-levels. According to the 1978 Waddell Report, "the O Level examination [was] aimed at the upper 20 percent of the full ability range and CSE catering for the next 40 per cent." Before the introduction of the CSE, the majority of schoolchildren left school without a nationally recognized qualification. CSE's were available in both academic and vocational subjects with five pass grades ranging from 1 (highest) to 5 (lowest). CSE grade 1 was equivalent to an O-level pass.

In 1988 the O-Level and CSE were replaced in the United Kingdom by the GCSEs (General Certificate of Secondary Education). The GCSE is an academically rigorous qualification typically taken over two years with studies starting at the beginning of Year 10 (ages 14-15) and final examinations taken at the end of Year 11 (ages 15-16).

III. Data

A. The UK Biobank

We will use data from the UK Biobank, a large, population-based prospective study initiated by the UK National Health Service (NHS) (Sudlow et al. 2015). Between 2006 and 2010, invitations were mailed to 9 million people between the ages of 40 and 69 who were registered with the NHS, which has contact details for an estimated 98% of the UK population. More than a half-million individuals responded and agreed to contribute data to the Biobank. The large sample size and the age range were chosen to allow a reliable assessment of the main determinants of common health conditions.

Participants were assessed at 22 centers throughout the UK. The assessment comprised a self-completed touch-screen questionnaire; a brief computer-assisted interview; physical and functional measures; and collection of blood, urine, and saliva. Information about health status, lifestyle, diet, psychosocial factors, and cognitive function were collected through the self-completed touchscreen questionnaire and the computer-assisted interview. The Biobank also collected a series of physical measures: blood pressure, heart rate, grip strength, anthropometrics, spirometry, bone density, arterial stiffness, eye examination, a hearing test, and a fitness test. 100,000 participants also wore accelerometers that recorded physical activity for 7 days. Every participant was genotyped. Finally, the Biobank has linked its data to extensive NHS health records with hospital-inpatient, hospital-outpatient, and primary-care data, which are unique because of the nearly universal health coverage through UK's NHS.

While these data will eventually allow for a rich analysis of these detailed and more reliable measures of health, for this paper, we will focus on two outcomes: body mass index (BMI) and self-reported diabetes. We chose these outcomes as a starting point for a variety of reasons. First, these variables were immediately available for the entire data set. Second, these two traits have been shown to be strongly associated in previous literature (Diabetes Prevention Program Research Group 2002, Ford et al. 1997, Narayan et al. 2007, Sinha et al. 2002), so

studying them jointly may shed further light on this relationship. Third, because BMI is a continuous measure, it will be possible to study the effect of the 1972 ROSLA on the distribution of BMI in addition to estimating the mean effects. Lastly, a very large genetic study of BMI (Locke et al 2015) has been performed, which allows us to generate a genetic predictor of BMI to study heterogeneous treatment effects by genetic risk.

Since we are studying the effect of the 1972 ROSLA, which only applied to individuals living in England and Wales at the time of the reform, we would optimally restrict our sample to those living in those countries at age 15. The UK Biobank only collects data on an individual's current residence and their country of birth. We therefore include only those individuals born in England and Wales as the best available proxy. Additionally, in order to reduce biases based on genetic ancestry, we restrict our analysis to those of European descent. A further discussion of this restriction is found in the following subsection on the genetic data.

Summary statistics for the variables relevant to this study are found in Table 1. Given that our analysis will estimate effects local to the treatment threshold (i.e. those born in a window of September 1, 1957), Panels A and B of this table give results for those born within 10 and 5 years of this date, respectively. In addition to a summary of the all of the individuals used in the analysis, we may be interested in a summary of these variable for those that were influenced by the reform (i.e. "compliers"). While these individuals can't be identified after treatment, we can restrict our sample to those dropping out of school at age 15 within 5 years prior to the ROSLA. These results are found in Panel C of Table 1.

From this table, we first notice that the variable for the age at which an individual left school is only available for part of the sample. This is because this question was only asked of the 65% of individuals who did not report having a college or university degree on a previous question. We will discuss the implications this has on our empirical strategy in the following sections.

Second, we note that the sample is slightly overweight, with an average BMI of roughly 27. To put this in context, a BMI in the range of 18.5 to 25 is considered healthy, a BMI in the range of 25 to 30 is considered overweight, and a BMI above 30 is considered obese. The sample of "compliers" has a mean BMI 1.2 points higher than that of the entire sample. At the time of data collection, which would be around age 60, the prevalence of diabetes reported in the sample is a little more than 3%, relative to a prevalence of more than 5% in the set of "compliers."

Last, our final sample is very large relative to previous literature. For instance, at the same bandwidth used to measure the first-stage effect of the 1972 ROSLA, we have approximately twice as many observations as Clark and Royer (2013). This will be very important since with samples as large as ours, we are much better powered to detect moderate mean effects and also are powered to estimate effects at various quantiles of the health distribution.

B. Genetic Data

The human genome is made up of strings of billions of molecules, or nucleotides, most of which are identical across the whole population. In approximately 10 million locations of the genome, however, an alternate nucleotide is found in more than 1% of the population. These locations are called single-nucleotide polymorphisms (SNPs) and the possible nucleotides observed at a SNP are called alleles. The alleles found at the various SNPs throughout the genome for an individual is called that individual's genotype.

Using saliva samples, each subject in the UK Biobank sample has been genotyped. This was done using the Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants and using the Affymetrix UK Biobank Axiom array for the remaining 450,000 participants.[2] This provided molecular genetic information at 641,018 markers throughout the genome. Using the known correlation structure of the human genome in European samples, the genotype was

[2] The SNPs genotyped by these two arrays overlap at over 95% of loci.

imputed for millions of common SNPs found in these populations. In the current release of the data, these full genotypic results are available for roughly one-quarter of the sample. The full release should be available by mid-2016.

As our measure of genetic risk for various outcomes, we will construct a polygenic score, which is defined as the weighted sum:

$$S_i = \sum_j g_{ij} w_j \quad (1)$$

where g_{ij} is the number of alleles for a given SNP j for individual i and w_j is the weight for SNP j. Previous work has shown that, although the explanatory power of each individual SNP is very small, a weighted sum of these SNPs with optimal weights could explain over 20% of the variation in educational attainment (EA) (Rietveld et al. 2013) and also over 20% of the variation in BMI (Locke et al 2015).

To construct polygenic scores, we combine the genetic data with SNP weights drawn from large genome-wide association studies (GWAS). To avoid over-fitting, we are using GWAS that do not include UK Biobank data when calculating SNP weights. Specifically, the weights for EA come from Okbay et al. (2016) and the weights for BMI come from Locke et al (2015). To further improve the predictive power of the polygenic scores, the weights taken from these large GWAS will be corrected for correlation using modern techniques based on Bayesian methods (Vilhjalmsson et al. 2015).

Interpretation of a polygenic score is not straightforward. They may be thought of as representing a set of causal pathways that begin with a person's genes and end with the trait of interest, passing through any number of biological and environmental steps along the way. For example, the polygenic score may capture direct pathways to BMI, such as changing a person's build. It may also simultaneously capture more complex pathways, such as altering how addictive nicotine is for a person, increasing their smoking if they grow up in an environment

where they are likely to start smoking, which may reduce their BMI but also may have epigenetic consequences, triggering many other biological mechanisms influencing BMI.

There are a couple threats to this causal interpretation. The common concern of reverse causality is not a problem in this case because an individual's genes are randomly assigned at conception conditional on the genes of that individual's parents. There are two forms of omitted variable bias, however, that may be concerning. First, if parental genes are a causal factor in their children's health, some bias may be introduced since the genotype of parents and children are correlated. We note, however, that studies that have looked at the association between within-family (i.e., across siblings) differences in genotypes and within-family differences in BMI find similar results to the individual level analyses, which are susceptible to these parental effects. This suggests that, if parental effects are present at all in the polygenic score, they are not likely to play a large role in the predictive power of the score relative to the causal factors.

Another form of potential omitted variable bias stems from differences in genetic ancestry. Over time, non-assortative mating between groups of different ancestries can lead to differences in the frequency of certain alleles within that group, leading to correlation between a person's genotype and their ancestry.[3] If groups with different ancestries tend to have differences in health for non-genetic reasons, a genetic score may be highly predictive of the outcome of interest, though this association is also not part of a biological causal pathway.[4]

To control for ancestry and avoid such bias, we will restrict our sample to those of European descent. In practice this this restriction will not significantly reduce our sample size since over 95% of our sample is of European descent. Additionally, we will include 15 principal components of the genetic data as controls. Since the primary information encoded in people's

[3] This type of correlation is called "stratification" in the genetics literature

[4] The canonical example of stratification was first presented in Hamer (2000). Imagine that a polygenic score was produced that was highly predictive of chopstick use. Our primary hypothesis would not be that these genes increased manual dexterity or taste for sushi, but rather that the score loads on alleles that are more common in Asian populations.

DNA is their ancestry, this has been shown to remove bias due to stratification almost entirely within European populations (Price et al. 2006).

Lastly, since one way in which stratification is manifested is geographically, we can evaluate the degree of stratification in our data by considering the spacial distribution of the genetic score. Panel A of Figure 1 plots the mean polygenic score for each cell of a grid over England and Wales. Each cell is color-coded according to its quartile among all of the cells displayed. For the most part, this map appears to suggest a uniform distribution of scores, with perhaps a slightly higher concentration of lower genetic scores in the South-East in the vicinity of London.

As a baseline comparison of what this map would look like when the polygenic score has no spacial correlation, we randomly generate a permutation of the score variables and generate the same map, which is found in Panel B of Figure 1. For Panel C, we first regress the score on a quadratic of latitude and longitude and a set of county fixed effects and plot the mean residuals of this regression. We see that this effectively removes any spacial correlation that may have been present in the raw scores. This highlights the need to verify that our results are robust to geographic controls and interactions if we'd like them to have a biological interpretation.

IV. Mean Effect of Education on Health

A. Empirical Strategy

We are first interested in estimating the mean effects of education on health. Due to biases introduced by reverse causality and omitted variables, we are unable to estimate this causal relationship by simple ordinary least-squares (OLS) estimation. Given that it is unlikely that parents were timing the birth of their children in response to the ROSLA that would go into effect 15 years later, we should be able to use the policy change as an instrument for educational attainment and calculate two-stage least squares (2SLS) estimates of the causal effects of an additional year of schooling at age 15 on health at the time of data collection.

Where possible, we follow the empirical strategy in Clark and Royer (2013) so our results are comparable to theirs. Specifically, we use a regression discontinuity (RD) design to estimate the first-stage and reduced-form effects of the 1972 ROSLA. This approach effectively compares the outcomes of those born immediately before September 1, 1957 and those born after, dividing the sample into plausible control and treatment groups. In practice, we use the specification described in Lee and Lemieux (2010). More precisely, we estimate parameters of the equation

$$Y_i = \beta_0 + \beta_1 After_i + f(DoB_i) + \boldsymbol{X}_i \boldsymbol{\beta}_4 + \varepsilon_i. \quad (2)$$

where Y_i is some outcome of interest for individual i (e.g., education or BMI); $After_i$ is one if the individual was born on or after September 1st 1957 (and 0 otherwise); DoB_i is the individual's date of birth; and the vector $\boldsymbol{X}_i$ contains a set of predetermined control variables such as gender.[5] The function $f(\cdot)$ captures birth-cohort trends in Y_i.

There are two standard ways to model this trend: a local linear approach and a global polynomial approach. In the local linear specification, $f(\cdot)$ is modeled as a linear function with different slopes before and after the discontinuity and observations distant from the discontinuity threshold are down-weighted. As may be inferred from its name, the global polynomial approach models $f(\cdot)$ as a polynomial that may differ on either side of the discontinuity as well. The polynomial is usually limited to second or third order to avoid overfitting. Gelman and Imbens (2014) and Imbens and Kalyanaraman (2012) suggest, however, that local linear specifications tend to be more stable. For this reason, the results presented in this paper are based on a local linear specification.[6] As is recommended in Imbens and Kalyanaraman (2012), we use a triangle kernel to weight the observations, and we use the optimal bandwidth rule proposed in that same paper to select the width of the kernel. We will additionally cluster the standard errors by month-of-birth.

[5] The inclusion of predetermined controls in equation (1) is not needed for identification but can improve the estimates' precision.
[6] We also report the global polynomial approach as a robustness check.

A key assumption of the RD framework is that the mean potential outcomes of Y_i—that is, the expected value of Y_i given an individual's treatment status—is continuous across the discontinuity. If this wasn't so, it would be impossible to disentangle the discontinuity in potential outcomes from the treatment effect. While it seems unlikely that such a discontinuity would arise in the population from one month to the next, this assumption may be violated if there is differential selection on either side of the threshold. We test for differential selection in two ways. First, we perform a McCrary Test (McCrary 2008), which tests for a discontinuity in the distribution of birth month over the threshold. More precisely, we estimate (2) where Y_i is a count of the number of individuals born in month i. Second, we test for balance in predetermined observable characteristics as suggested by Lee and Lemieux (2010). In this test, we estimate (2), where Y_i is some outcome determined before the ROSLA was implemented. In our case, we use an individual's coordinates of birth and his or her genetic score for BMI and for educational attainment.

In order to estimate the first-stage regression, we need a measure of the educational attainment of the individuals in the sample. The UK Biobank asked the question "At what age did you leave school?" though as explained in the previous section, this was only asked of the 65% of respondents who reported not having graduated from college or university. Since the direct impact of the ROSLA was on those dropping out of school at age 15, a simple way to resolve this data limitation is to code individuals who have a college degree as having left school at some age greater than 16 and to estimate (2) for a binary outcome variable which indicates whether an individual was in school till at least age 16.

This approach, however, may yield biased 2SLS estimates. While a simple human capital model would predict that there would only be a response in the number of students in school till age 16, a model with signaling (Lang and Kropp 1986), spillovers (Acemoglu and Angrist 2000), or ability learning may yield increases in the fraction of students remaining in school till many other ages. We can estimate the magnitude of the bias by performing RD

regressions for binary outcomes indicating whether a student was still in school till various age thresholds other than age 16. Then the 2SLS will be inflated by a factor of

$$\frac{\sum_j \beta_{1,j}}{\beta_{1,16}}, \tag{3}$$

where $\beta_{1,i}$ is the $After_i$ coefficient from (2) corresponding to a specification where the outcome variable is an indicator of whether the individual is in school till age j.

B. Results

Scatter plots corresponding to the McCrary and balance tests are found in Figures 2 and 3. Numerical results of these tests are contained in Table 2. These tests provide no evidence that there is any differential selection across the discontinuity threshold. In each case the point estimate is insignificant and the standard errors are precise enough to rule out even small mean differences. This reinforces our confidence that the assumptions underlying our empirical strategy are met, and we therefore continue with our analysis.

Figure 4 shows a scatter plot of the fraction of individuals still in school binned by six-month intervals. Figure 5 plots the corresponding discontinuity coefficients. Table 3 displays the results for various specifications of (2).

These results show that the bulk of the increase in educational attainment was in the fraction remaining in school till age 16, increasing by around 15 percentage points.[7] There also appears to have been a small, though significant, increase in students remaining in school till age 15. It is possible that this is due to stricter enforcement of existing compulsory schooling laws after the ROSLA Order. This effect would also be present if there was a small amount of misreporting. There also appears to be small, statistically-insignificant increases in the fraction of students still in school at ages 17 and 18.[8] So it does not appear that the ROSLA led to large

[7] These results are qualitatively similar to what is seen in the Health Survey for England, as reported in Clark and Royer (2013).
[8] We don't test other ages in our data since some individuals may be graduating college at older ages.

increases at ages other than age 16. If we use the point estimates at all thresholds tested as the true impact at each age, these results suggest that our IV results may be inflated by about 5%.

Table 4 and Figure 6 display the impact of the ROSLA on mean BMI. The first two columns show the OLS relationship between BMI and remaining in school till age 16. The two middle columns show reduced-form estimates. The last two columns show 2SLS estimates, where we use the discontinuity to instrument for remaining in school till age 16. Columns 1, 3, and 5 show results without controls while columns 2, 4, and 6 show results that include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth.

The reduced-form suggests that there was a small decrease in BMI of 0.06 BMI points, though our results are not precise enough to rule out a zero effect from the 95% confidence interval. Table 4 also reports the 2SLS estimates of the impact of an additional year of education on BMI. It implies that the additional year of education at 16 lead to a decrease in BMI of approximately 0.5 BMI points, though this is unsurprisingly also insignificant. This is much smaller in magnitude than the OLS estimate of a decrease of 1 BMI point, however the standard errors on the 2SLS estimate are large enough that we can't reject that these estimates are the same.

The results for diabetes, which are found in Figure 7 and Table 5, are stronger. In this case, we estimate that the ROSLA decreased the prevalence of diabetes by more than 0.3 percentage points. We also estimate that the effect of an additional year of school on diabetes risk is a decrease of almost 3 percentage points, roughly double the OLS estimate. In the specification with controls, these estimates are significant at the 1% level. We note that that in the specifications without controls, both the reduced form and 2SLS estimates have a t-statistic of almost exactly two. As a result, we present these results with caution.

V. Heterogeneous Effects

A. Empirical Strategy

An advantage of continuous measures of health such as BMI is that we are able to estimate how education affects different points of the BMI distribution rather than just the effect on the mean. We will do this in a number of ways. First, we will estimate a "reduced-form" quantile model, which will estimate the impact of the ROSLA on different quantiles of the BMI distribution for all individuals in the data. Second, as an "instrumental variable" counterpart to quantile estimation, we will implement a method by Frandsen, Frölich, and Melly (2012), which gives quantile treatment effects of the impact of an additional year of education on various quantiles of the BMI distribution *for compliers*. Lastly, we will create binary measures of BMI at several thresholds and estimate how an additional year of education affects the probability that an individual has a BMI above that threshold. Details of these approaches are found in the following paragraphs.

Quantile estimation, introduced by Koenker and Bassett (1978), generates coefficients for the model

$$P(Y_i \leq \beta_0 + \beta_1 After_i + f(DoB_i) + X_i\boldsymbol{\beta}_4 \mid After_i, DoB_i, X_i) = \tau \quad (4)$$

for each quantile τ, where the variables above are defined as in (2). This approach defines a function $\beta_1(\tau)$, which represents how much quantile τ increases for those born after September 1, 1957. More details can be found in Hao and Naiman (2007).

Interpretation of the results of this model is not as straightforward as with mean effects. Specifically, this approach only estimates how the τ*-th quantile* is affected as a result of the policy change, in contrast to estimating how much the BMI *of the individual* at the τ-th quantile is affected as a result of the policy change. As a result, output of this estimator should be thought of as measuring the effect on the population as a whole rather than individuals.

If we would like to treat quantile estimates as "the effect on the individuals at each quantile," we would need an additional strong assumption of "rank invariance." This assumption

requires that the rank order of individuals remains the same before and after treatment. More precisely, for any two individuals, A and B, A would have a higher BMI than B when they both leave school at age 15 if and only if A would also have a higher BMI than B when they leave school at age 16. In the following section on genetic heterogeneity, we will show that this assumption is likely false since individuals around the same quantile of the BMI distribution appear to respond very differently to an extra year of schooling based on their genetic risk for obesity.

In a traditional instrumental variable setting, it is assumed that the outcome for always-takers (i.e. individuals who would remain in school until at least age 16 even without the ROSLA) wouldn't change after the policy change. Therefore, you can estimate the effect on compliers by rescaling the average effect on the whole population by the fraction of compliers (i.e. individuals who would drop out at age 15 before the ROSLA but would be constrained to stay in school an additional year if they were part of the treated group). Since the distribution of BMI is possibly different for compliers than it is for always-takers, some quantile of BMI in the entire population may not correspond to the same quantile of BMI in the population of compliers. As a result, to estimate the effect on a quantile of the distribution of compliers, you must also transform the quantiles of the "reduced form" estimates so they correspond to the quantiles of the complier distribution.

To do this, we follow the approach proposed by Frandsen, Frölich, and Melly (2012), which is a variation on a Wald Estimator that estimates quantile treatment effects in an RD setting. This approach relies on four assumptions:

1. The probability of treatment changes discontinuously at the RD threshold,
2. The distribution of potential outcomes is continuous across the RD threshold,
3. No one who would be treated before the RD threshold is untreated after it, and
4. There is positive density at the RD threshold.

These assumptions are nearly identical to those of the standard RD design. The sole difference is assumption 2 which normally only requires that the expectation of potential outcomes be continuous across the RD threshold. Though this assumption is slightly stronger, the timing of the implementation of the ROSLA makes us think it is unlikely that any quantiles of the potential outcome distribution would discontinuously shift at the RD threshold. This is again supported by the smoothness of predetermined characteristics examined in the previous section.

As a final specification, we will choose some BMI threshold t, and estimate the reduced-form equation (2) where the outcome variable is an indicator of whether the individual has a BMI above t. We can then estimate by 2SLS the impact of a year of education on the probability that the individual has a BMI above t.

We note that this specification is closely related to the quantile regression approach described above. Both approaches effectively estimate the difference in the cumulative distribution function of potential outcomes of compliers in the treated and untreated state. Quantile regression holds the percentile fixed and takes the horizontal difference of BMI while the secondary binary approach holds the BMI fixed and takes a vertical difference of percentile.

B. Results

Figure 8 shows a scatter plot of how the 25th, 50th, 75th, and 90th percentiles of the whole population vary grouping individuals in 6-month bins. These figures suggest that there is no discernible discontinuous change in these quantiles after the ROSLA at the 25th percentile and the median, though some effects may be seen at the upper quantiles. Figure 9 plots the estimated reduced-form discontinuity at a number of quantiles between the 10th and 90th. This figure also includes a dotted line that shows the quantile function of BMI before the reform.

As can be seen, no effect significantly distinct from zero can be detected below the 40th percentile, which are quantiles corresponding to healthy weights (i.e. a BMI below 25). As we move up the distribution of BMI from this point, however, the estimated effect of the ROSLA on

this portion on the distribution begins to peel away from the x-axis, achieving statistical significance for quantiles above the 70th. This means that the ROSLA is estimated to have had the overall effect of shifting the distribution from a less health to a more healthy range while having at most a minor impact on the health portion of the BMI distribution. Additionally, it appears that the distribution shifted most at quantiles corresponding to the least healthy levels of BMI.

Figure 10 displays a similar figure corresponding to the IV estimates of the effect of an additional year of school on the BMI distribution of compliers. As expected, this figure has roughly the same shape, deviating from the x-axis at the same quantile corresponding to the transition from a healthy weight to overweight. One difference between Figure 10 and Figure 9 is that the point of departure in the IV figure arrives at an earlier quantile than in the reduced-form figure. This is because the group of always-takers, which are differenced out by this approach, tend to have a lower BMI on average. As a result, the 40^{th} percentile of the whole distribution corresponds to the 20^{th} percentile of the distribution of compliers. These effects are large, suggesting that an additional year of education decreases the 90^{th} percentile of the BMI distribution by 2 BMI points from a baseline of 35.

Finally, Figure 11 shows the results for the series of RD specifications corresponding to various BMI thresholds and Figure 12 shows the corresponding IV estimates. As with the quantile regressions, these estimates suggest that the ROSLA had little impact on the BMI of individuals with an untreated potential outcome in a healthy range, but as we estimate the change in probability that a person has a BMI above thresholds in an unhealthy range, we are able to detect significant improvements in health. Based on the 2SLS coefficients, these improvements are as large as 9 percentage points at a BMI threshold of 30, from a baseline of 33%.

In contrast to the quantile regressions, where the discontinuities and BMI quantile had an approximately monotonic relationship, the effect at higher BMI levels seems to fall after a

threshold of about 30. This is primarily driven by boundary effects. As we move to higher and higher BMI thresholds, the fraction of individuals with a BMI above that threshold gets smaller and smaller. If we would like to scale these coefficients so they are comparable across thresholds, we can divide them by the baseline fraction of individuals with a BMI above that threshold. Then the coefficients would be an estimate of the *percent* decrease in probability that an individual would have a BMI above a certain level as opposed to a *percentage point* decrease. These scaled reduced-form results are found in Figure 13. As can be seen, the trend of these coefficients more closely matches the patterns seen in the quantile regressions, suggesting that the ROSLA decreased the fraction of individuals with a BMI above each threshold over 30 by about 5%.

VI. Genetic Heterogeneity

A. Empirical Strategy

In the previous section, we showed evidence that there is a significant amount of heterogeneity in the treatment effects of education on BMI. We may be interested in if this heterogeneity varies systematically by observable characteristics, such as an individual's genetic risk for poor health. Genes are a particularly interesting trait to use in this context because they are assigned at birth and are immutable. In this section, we employ two related approaches: one using a parametric continuous specification and the other using a more nonparametric categorical specification.

In both cases, we will estimate the following first-stage and reduced-form models

$$Y_i = \boldsymbol{G}_i\boldsymbol{\beta}_0 + After_i \times \boldsymbol{G}_i\boldsymbol{\beta}_g + f(DoB_i) + \boldsymbol{X}_i\boldsymbol{\beta}_4 + \varepsilon_i \quad (5)$$

where Y_i, $After_i$ and DoB_i are defined as before. In the continuous specification, $\boldsymbol{G}_i$ is a vector consisting of a constant, an indicator of whether an individual has genetic data, and the genetic score for BMI described in section 3. This way, the element of $\boldsymbol{\beta}_g$ associated with the score will measure how much an increase in the score is associated with an increase in the treatment

effect. More precisely, the effect of the ROSLA on outcome Y_i for an individual with genetic score S_i would be

$$\beta_{g,1} + \beta_{g,3} S_i \tag{4}$$

So in this case, we are just-identified for our 2SLS estimates, using the vector $After_i \times \boldsymbol{G}_i$ to instrument for the endogenous vector $Edu_i \times \boldsymbol{G}_i$, where Edu_i is the binary variable of whether individual i remained in school till age 16.

In the categorical case, we first divide the sample into 4 groups: those without genetic data and three terciles based on each genotyped individual's genetic BMI score. We then estimate (3), where $\boldsymbol{G}_i$ is a binary vector indicating which group the individual is in. In this specification, the elements of the parameter $\boldsymbol{\beta}_g$ may be interpreted as the effect of the ROSLA on the outcome Y_i for each group. Similarly, the corresponding 2SLS estimates may be interpreted as the effect of remaining in school till age 16 on the outcome.

In each case, we may be worried that any results found may not be driven by biology but instead represent geographic variation in the score. To test this, we generate a set of rich geographic controls including a quadratic in latitude and longitude and county fixed effects. We then estimate (3), including an interaction of these controls with the $After_i$ variable. If the heterogeneity results are driven primarily by geographic variation in the genetic score, then we'd expect to see large attenuation in the interaction effects.

B. Results

First-stage results for the continuous and categorical specification are found in Table 6. A visual representation of the categorical results is found in Figure 14. We see here that there is a strong amount of heterogeneity in the degree of treatment, where each standard deviation increase in the BMI genetic score is associated with an increased first-stage effect of 2.3 percentage points. In the categorical case, we see a highly significant 5 percentage point difference between those in the lowest BMI score tercile and those in the highest tercile. These

relationships hold up in all specifications, including those where the discontinuity variable is interacted with the geographic controls.

Comparable results for the reduced form specifications for BMI are found in Table 7 and Figure 15. We note that in the categorical specification, we estimate that the ROSLA decreased mean BMI of those with a low BMI score by 0.2 BMI points from a baseline of 26.1, while there was no real impact on those in the other BMI terciles. This is particularly striking given the first-stage results described in the previous paragraph—the group that was most strongly treated by the ROSLA saw the smallest reductions in BMI.

We may be interested if this relationship holds through the distribution as opposed to just at the mean. Remembering that the largest effects were seen in the highest quantiles of the BMI distribution, we estimate the same specifications as above for the effect of the ROSLA on the 90 percentile of BMI conditional on the BMI score. These results are found in Table 8 and Figure 16. We note that the same patterns hold in these regressions, though the point estimates tend to be roughly three to four times larger. As before, the largest effects on the 90^{th} percentile of BMI are found in those with a low genetic score for BMI.

Table 9 and Figure 17 contain the comparable results for diabetes risk. In this case, we see strongly significant main effects and heterogeneous effects. As before, these results are robust to the inclusion of geographic controls interacted with the discontinuity variable. In contrast to the results for BMI, however, we see the strongest effects in those in the highest BMI score tercile, with a reduction of 1 percentage point from a baseline of 5% and little reduction in the group with the lowest BMI scores.

Two-stage least squares results are found in Table 10 and Figures 18 and 19. The patterns are similar to those seen in the reduced form specifications. These estimates suggest that staying in school till age 16 as opposed to dropping out at 15 leads to a reduction in BMI of 1.5 BMI points in the group of individuals with the lowest BMI scores while little to no effect is scene in the group with the highest BMI scores. In contrast, those with the highest BMI score

experience a decrease of around 6 percentage points in diabetes risk for the additional year of school relative to insignificant decreases in those with the lowest score.

These results are surprisingly strong, which may call into question whether these results are primarily driven by chance. Fortunately, these estimates are based only on the 25% of the observations for which we currently have genetic data. The complete genetic data will become available within the next several months, which will allow us to verify if these results replicate in a sample three times as large as this one.

C. Potential Biological Mechanisms

As discussed in Section 3, these genetic heterogeneity results may be hard to interpret since the genetic score represents any number of unknown causal pathways. In this section, we attempt beginning to untangle some of these pathways and propose other tests that may elucidate the mechanisms driving the genetic heterogeneity observed.

Given that the results described above are robust to the inclusion of rich geographic controls, we take this as evidence that the heterogeneity is driven primarily by biological processes. To understand these processes, we first turn to the literature to try to understand what is known about the regions of the genome where the polygenic score puts the most weight. Locke et al (2015) and Finucane et al (2015) use the GWAS results on which our BMI genetic score weights are based to separate out the primary sources of the signal. In both cases, they find that the strongest effects detected are concentrated in regions of the genome more strongly associated with the central nervous system. Locke et al further indicate that the strongest systems affected are those associated with central appetite regulation, learning, cognition, emotion and memory. On the other hand, they test for a concentration of signal in areas associated with the digestive, endocrine, and musculoskeletal systems and find nothing.

In addition to being interested in the regions of the genome where the signal of the score is concentrated, we may be interested in how scores for different traits may be related to the genetic score for BMI. While this does not reveal any causal relationship between the two traits,

it may highlight the degree to which two traits share similar causal pathways. As our measure of this relationship, we use the genetic correlation, which signifies the correlation between the scores for a pair of traits if the weights for the score were estimated in an infinite sample. We draw these estimates from Bulik-Sullivan et al (2015), which reports the estimated genetic correlation for a large set of traits including BMI. Selected results are found in Table 11.[9] Of the 42 traits compared to BMI, 27 have a Bonferroni corrected p-value reaching statistical significance at the 5% level. The strongest correlations are found between BMI and other related measures of obesity, all having a correlation of nearly one. Diabetes and related traits (such as fasting insulin) are also strongly related with correlations around 0.5. At the next tier, traits related to educational attainment have a genetic correlation of about -0.3 and traits related to smoking behavior have a genetic correlation of 0.3. Other traits tested, including a number of mental and physical health outcomes, are estimated to have genetic correlations smaller in magnitude than that.

Though genetic correlation provides insight into which traits are genetically related to BMI, this relationship is relevant for the direct effect of genes on the traits. We are instead interested in the common causal pathways related to the genetic heterogeneity of the effect of education on health. To estimate the overlap related to the interaction of the ROSLA and genetic risk, we create a genetic score for educational attainment based on the GWAS results of Okbay et al (2016), regress the genetic score for BMI on the genetic score for educational attainment, and estimate (3) using the residual of this regression in the place of the raw genetic score for BMI. The resulting coefficient from this specification will give an estimate of the

[9] Many of the traits contained in Bulik-Sullivan et al (2015) are related to one another. In selecting which traits to include in Table 11, we tried to only omit a trait if a close proxy was left in the remaining traits.

contribution of the BMI score to treatment effect heterogeneity that is not captured by the pathways represented by the genetic score for educational attainment.[10]

The results from this approach are found in Table 12. As we can see here, the coefficient corresponding to the residual of the BMI score falls by about 25% in the first-stage regressions and by about 20% in the diabetes regressions, suggesting that the majority of the sources of this heterogeneity are from pathways unique to the BMI score. There is very little difference, however, in the BMI regression coefficient when we use the residual of the BMI score. We highlight, however, that the amount of information we can infer from this analysis is limited. While decomposing the signal into "the fraction relevant to a genetic score for educational attainment" and its residual may be interesting in refining our hypotheses of the source of the heterogeneity observed, since we know little about the mechanisms that each score represents, forming strong conclusions from these results is not possible.

A potential future line of analysis would be to skip the biological mechanisms, which researchers are still trying to understand and for which we have little data, and consider more distal steps in the causal chain. For instance, we can estimate the impact of the ROSLA on factors that we believe may be on the causal pathway for BMI, including earnings, occupation, caloric intake, and preventative health care. This approach will generate not causal estimates of the pathway (i.e. we won't be able to distinguish between education causing increased earnings causing improved health and education causing improved health causing increased earnings), but we should be able to put bounds on which relationships are most closely related.

VII. Robustness

A. Bandwidth analysis

The bandwidths used in the analyses of this paper are based on the optimal bandwidth algorithms proposed in Imbens and Kalyanaraman (2012). There are two competing forces at

[10] Educational attainment was used in this case because it is a trait related to this study where weights have been estimated in a large sample. In the future, a similar type of analysis can performed using scores for a variety of other traits.

play in this parameter tuning problem, often referred to as the bias-variance trade-off. First, the wider the bandwidth, the more data are included, increasing the precision of the estimate. Second, the narrower the bandwidth, the less curvature there is in the running variable, which generates bias in the local linear model due to misspecification.

In order to confirm that our results are robust to differences in bandwidth, we estimate the first-stage and reduced form models at a number of bandwidth between 0.5 and 1.5 times the optimal bandwidth. The coefficients associated with the discontinuity are found in Table 13 and plotted in Figures 20 through 23. As can be seen, the magnitude of the effect is quite stable at all bandwidth considered, though the confidence intervals corresponding to wider bandwidths are tighter, as expected. Considering how linear the relationship is between month of birth and the outcomes we consider, seen in Figures 4, 6, 7, and 8, it is unsurprising that the point-estimate varies little by bandwidth.

B. Local Linear versus Global Polynomial Specifications

While local linear specifications are often more stable than polynomial specifications, if there is curvature in the trend leading up to the discontinuity threshold, these estimates may still be substantially biased. While there appear to be limited local curvature in the scatter plots presented, we also perform a global polynomial specification, which includes all available data but allows for cubic trends that can differ before and after the discontinuity. We expect these estimates to be similar to those of the local linear specification. A comparison of the RD and 2SLS estimates by both the local linear and global polynomial approaches is found in Table 14.

In each case, the estimates based on a local linear and a global polynomial approach are statistically indistinguishable. We note, however, that the estimate for the global polynomial approach is approximately twice as large as the local linear approach for the reduced-from and 2SLS specifications when the outcome is diabetes. Similarly, the quantile regression estimate for 90th percentile BMI is 32% larger in the global polynomial approach than in the local linear. Since our preferred specification (local linear) is the more conservative in each of these cases,

we are not concerned by this difference. As a whole, we take these results as evidence that our results are at worst not driven by the specification of the trend, and they are possibly conservative estimates of the true effect.

VIII. Discussion

In this paper, we have identified several key empirical results. First, we find marginally significant evidence that the ROSLA decreased the diabetes risk of those affected by the reform. We also estimate that the reform decreased mean BMI in our sample, though this result is not statistically significant. By comparison, Clark and Royer (2013) find no significant effect for any of the health measures they considered, including BMI (they do not have a measure of diabetes).

There are many possible explanations for why we detect an impact of the ROSLA on diabetes when Clark and Royer find no impact on any measure. Our primary hypothesis is that, with the exception of their data used for estimating the effect of education on mortality, their sample is roughly half the size of ours. As a result, their standard errors generally cannot rule out moderate effect sizes. For instance, their confidence interval of the impact of education on BMI comfortably contains our (insignificant) estimate of -0.53 BMI points.

It may seem surprising that even with such a large sample size, we are not able to detect convincingly significant effects of the 1972 ROSLA. Part of the reason for this is a limitation of the study design. Since the standard errors are clustered by month of birth, and since the treatment variable does not vary within a cluster, there is a strictly positive lower-bound on the size of the standard errors for a fixed number of clusters even with an arbitrarily large sample. In fact, since the optimal bandwidth is decreasing in sample size, there may be a point at which the increase in precision from adding addition observations is dominated by the loss in precision by removing clusters. Since the sample size in this study is so large, we anticipate that further increases in sample size would not further increase the precision of estimates of mean effects in an RD framework.

This does not mean that an RD framework to study the mean effects of the ROSLA on health will be fruitless going forward. The magnitude of the standard errors are not only a function of the sample size and number of clusters but also of the residual variance. If subsamples are identified with less residual variance or if additional controls are added to absorb what is left of the residual variance, this could have the same effect as adding additional samples. For instance, the genetic score for BMI currently explains nearly 7% of the variance in BMI. Improving these genetic predictors and finding other predictors of comparable explanatory power are likely the best strategies for estimating these parameters with sufficient precision.

As a second important empirical result, we find a significant amount of heterogeneity in the effect of education on different quantiles of the BMI distribution for compliers. This means that although the mean effect of education on BMI may be moderate, there are portions of the population for whom the effect is large. Conveniently, these large effects are concentrated in the high end of the BMI distribution, leading to a net compression of the distribution towards a more healthy range. This additionally may explain why the mean effect is difficult to detect: in the half of the population with BMI in a healthy range, education doesn't appear to decrease.

Third, using genetic data we are able to identify individuals for whom the effect of education on health is strongest. Oddly, the data suggest that those at high genetic risk for obesity tend to reduce their BMI least with an additional year of education relative to those with a low genetic risk for obesity. This seems to directly contradict the heterogeneity results which suggest that those with a high potential BMI absent treatment respond most to an additional year of schooling. These two results can be rationalized, however, with the understanding that there are many factors that influence BMI, including those with genetic and non-genetic roots. If these results hold up in the soon-available replication sample, these results simply suggest that an additional year of schooling more easily mitigates the non-genetic factors that lead to a high BMI.

In contrast to the genetic heterogeneity results for BMI, the heterogeneity for diabetes goes in the opposite direction: those at high genetic risk for BMI see the largest reductions in diabetes risk relative to those with low genetic risk for BMI. This may partially be due to boundary effects since on 2.5% of individuals with a low BMI score report having diabetes while 4.8% of high BMI score individuals report it. This result, however, does highlight that BMI and diabetes are not as strongly related as expected; the largest decreases in diabetes are seen in a group for which there is no comparable decrease in BMI. Future work is needed to understand what behaviors may be changing (e.g. diet or exercise) that play a more significant role.

IX. Conclusion

In this paper, we have presented evidence that the effects of education on health are strongly heterogeneous through the population. This may explain why even in our data with sample of unprecedented size, along with several previous studies of the impact of education on health, have had difficulty detecting significant effects. This result may be important for policy makers who need to project the potential welfare impacts of reforms that increase the educational attainment of the population. While the mean effects on long-term health may only be moderate, large effects can be seen is certain segments of the health distribution, in this case as measure by BMI.

It is also of interest that some of the heterogeneity can be predicted by observable characteristics, in this case genotype. While it is not feasible to expect targeted interventions meant to improve health by increasing the educational attainment of some groups based on their genotype, this research is a good proof-of-concept that similar approaches could very likely be successful in other settings. For instance, smokers may choose to use genetic information to inform themselves about how successful various expensive pharmaceutical interventions may be meant to help them quit.

This paper makes a very simple first attempt at understanding the biological mechanisms that drive the observed heterogeneity. This analysis suggests that the BMI score is

primarily driven by regions of the genome related to cognitive function, such as central appetite regulation, learning, cognition, emotion and memory. We also found that, while some of the heterogeneity was driven by variation common to a genetic predictor of educational attainment, the majority of the heterogeneity is driven by variation independent of the EA score. More careful work to understand these mechanisms should be conducted.

This paper leaves many question open for future research. First, some of the results reported are only marginally significant. Over the next few months as the complete genetic data of the UK Biobank are made available, we will be able to check if these results are replicated in the full sample. Second, as more precise GWAS estimates become available and tools for creating more predictive genetic scores are developed, it may be possible to produce sufficiently precise estimates of the mean effects of education on health that have thus far eluded research. Third, this paper does not take full advantage of the rich variable available in the UK Biobank. In the future, we will use several of the continuous, direct measures of health, such as spirometry tests, blood pressure, and glucose levels. We hope that these variables will lead to more precise estimates since they should be less susceptible to error due to misreporting and will allows us to estimate effects on the distribution of health as we did with BMI in the paper.

References

Acemoglu, D., & J. Angrist (2000). “How large are the external returns to schooling? Evidence from compulsory schooling laws.” *NBER Macroecon Annu*, *2000*, 9-59.

Albouya, Valerie & L. Lequien (2009). “Does compulsory education lower mortality?” *Journal of Health Economics*, 28,155–168.

Barsky, R. B., M.S. Kimball, F.T. Juster, & M.D. Shapiro (1995). “Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement survey.” NBER Working Paper 5213.

Becker, G. S., & C.B. Mulligan (1997). “The endogenous determination of time preference.” The *Quarterly Journal of Economics*, 729-758.

Bierut, L.J., J.A. Stitzel, J.C. Wang, A.L. Hinrichs, R.A. Grucza, X. Xuei, ... & W.J. Horton (2008). “Variants in nicotinic receptors and risk for nicotine dependence.” *American Journal of Psychiatry*, 165(9), 1163-1171.

Bitler, M.P., J.B. Gelbach, & H.W. Hoynes (2006). “Welfare reform and children's living arrangements.” *Journal of Human Resources*, *41*(1), 1-27.

Bulik-Sullivan, B., H.K. Finucane, V. Anttila, A. Gusev, F.R. Day, P.R. Loh, ... & M.J. Daly (2015). “An atlas of genetic correlations across human diseases and traits.” *Nature genetics*, 47, 1236-1241.

Carneiro, P., J.J. Heckman, & E.J. Vytlacil (2010). “Estimating marginal returns to education.” NBER Working Paper 16474.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuder, N. Turner, ... & D. Cutler (2016) “The association between income and life expectancy in the United States, 2001-2014.” *Journal of the American Medical Association*. Published online April 10, 2016. doi:10.1001/jama.2016.4226.

Clark, D. & H. Royer (2013). “The Effect of Education on Adult Mortality and Health: Evidence from Britain.” *American Economic Review*, 103(6), 2087–2120.

Cutler, D.M., A. Lleras-Muney, & T. Vogl (2008). "Socioeconomic Status and Health: Dimensions and Mechanisms." NBER Working Paper 14333. http://www.nber.org/papers/w14333

Cutler, D.M., & A. Lleras-Muney (2010). "Understanding differences in health behaviors by education." *Journal of Health Economics*, 29(1), 1-28.

Cutler, D.M., & A. Lleras-Muney (2010). "The education gradient in old age disability." *Research Findings in the Economics of Aging*. Chicago: University of Chicago Press, 101-120.

Cutler, D. M., & A. Lleras-Muney (2012). “Education and health: insights from international comparisons.” NBER Working Paper 17738.

Diabetes Prevention Program Research Group (2002). “Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.” *The New England Journal of Medicine*, *346*(6), 393.

Finucane, H.K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.R. Loh, ... & S. Ripke (2015). “Partitioning heritability by functional annotation using genome-wide association summary statistics.” *Nature genetics*, *47*(11), 1228-1235.

Ford, E.S., D.F. Williamson, & S. Liu (1997). “Weight change and diabetes incidence: findings from a national cohort of US adults.” *American Journal of Epidemiology*, *146*(3), 214-222.

Frandsen, B. R., M. Frölich, & B. Melly (2012). “Quantile treatment effects in the regression discontinuity design.” *Journal of Econometrics*, *168*(2), 382-395.

Gelman, A., & G. Imbens (2014). “Why high-order polynomials should not be used in regression discontinuity designs” NBER Working Paper 20405.

Goldman, D.P., & J.P. Smith (2002). “Can patient self-management help explain the SES health gradient?” *Proceedings of the National Academy of Sciences*, 99(16), 10929-10934.

Grenet, J. (2013). “Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence from French and British Compulsory Schooling Laws.” *The Scandinavian Journal of Economics*, 115(1), 176-210.

Grossman, M. (2015). *“*The relationship between health and schooling: What’s new?” NBER Working Paper 21609.

Hamer, D.H. (2000). “Beware the chopsticks gene.” *Molecular Psychiatry*, *5*(1), 11-13.

Hao, L., & D.Q. Naiman (2007). *Quantile regression* (No. 149). Sage.

Imbens, G., & K. Kalyanaraman (2012). “Optimal bandwidth choice for the regression discontinuity estimator.” *The Review of Economic Studies*, 79(3), 933-959.

Jürges, H., K. Eberhard & S. Reinhold (2013). “The effect of compulsory schooling on health—evidence from biomarkers.” *Journal of Population Economics*, 26, 645–672.

Koenker, R., & G. Bassett Jr (1978). “Regression quantiles.” *Econometrica: Journal of the Econometric Society*, 33-50.

Kemptner, D., J. Hendrik & S. Reinhold (2011). “Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany.” *Journal of Health Economics*, 30, 340–354.

Lang, K., & D. Kropp (1986). “Human capital versus sorting: the effects of compulsory attendance laws.” *The Quarterly Journal of Economics*, 609-624.

Lee, D.S., and T. Lemieux (2010). "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2), 281-355.

Lleras-Muney, A. (2005). “The relationship between education and adult mortality in the United States.” *The Review of Economic Studies*, *72*(1), 189-221.

Locke, A., B. Kahali, S.I. Berndt, A.E. Justice, T.H. Pers, F.R. Day, … & D.C. Croteau-Chonka (2015). “Genetic studies of body mass index yield new insights for obesity biology.” *Nature*, 518(7538), 197-206.

Marmot, M. G., G. Rose, M. Shipley, & P.J. Hamilton (1978). “Employment grade and coronary heart disease in British civil servants.” *Journal of Epidemiology and Community Health*, 32(4), 244-249.

Marmot, M.G., S. Stansfeld, C. Patel, F. North, J. Head, I. White, ... & G.D. Smith (1991). “Health inequalities among British civil servants: the Whitehall II study.” *The Lancet*, 337(8754), 1387-1393.

McCrary, J. (2008). “Manipulation of the running variable in the regression discontinuity design: A density test.” *Journal of Econometrics*, *142*(2), 698-714.

Narayan, K.V., J.P. Boyle, T.J. Thompson, E.W. Gregg, & D.F. Williamson (2007). “Effect of BMI on lifetime risk for diabetes in the US.” *Diabetes Care*, 30(6), 1562-1566.

Okbay A., J.P. Beauchamp, M.A. Fontana, J.J. Lee, T.H. Pers, C.A. Rietveld, … D.J. Benjamin (2016). “Genetic variants associated with education are linked to brain function and disorders.” *Nature*, forthcoming.

Oreopoulos, P. (2006). “Estimating average and local average treatment effects of education when compulsory schooling laws really matter.” *The American Economic Review*, 152-175.

Oreopoulos, P. (2007). “Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling.” *Journal of Public Economics*, 91(11–12), 2213–2229.

Perez-Arce, F. (2011). “The effect of education on time preferences.” RAND Working Paper WR844. http://www.rand.org/pubs/working_papers/WR844.html

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, & D. Reich (2006). “Principal components analysis corrects for stratification in genome-wide association studies.” *Nature Genetics*, *38*(8), 904-909.

Powdthave, N. (2010). “Does Education Reduce the Risk of Hypertension? Estimating the Biomarker Effect of Compulsory Schooling in England.” *Journal of Human Capital*. 4(2), 173-202.

Rietveld, C.A., S.E. Medland, J. Derringer, J. Yang, T. Esko, N.W. Martin, … & E. Albrecht (2013). "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment." *Science* 340(6139), 1467-1471.

Silles, M.A. (2009). “The causal effect of education on health: Evidence from the United Kingdom.” *Economics of Education Review*, 28, 122–128.

Sinha, R., G. Fisch, B. Teague, W.V. Tamborlane, B. Banyas, K. Allen, ... & R.S. Sherwin (2002). “Prevalence of impaired glucose tolerance among children and adolescents with marked obesity.” *New England Journal of Medicine*, *346*(11), 802-810.

Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, … & R. Collins (2015) “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.” *PLoS Med* 12(3), e1001779.

Vilhjalmsson, B., J. Yang, H.K. Finucane, A. Gusev, S. Lindstrom, S. Ripke, … & B. Liu (2015). "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores." *American Journal of Human Genetics*, 97(4), 576-592.

Waddell, J. (1978). "The Waddell Report - School Examinations." *Education in England*, London: Her Majesty's Stationery Office.

De Walque, D. (2007). “Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education.” *Journal of Health Economics*, 26(5), 877-895.

Tables

Table 1: Summary Statistics

	(A) Within 10 Years					(B) Within 5 Years					(C) Compliers				
	N	Mean	St Dev	Min	Max	N	Mean	St Dev	Min	Max	N	Mean	St Dev	Min	Max
Education Variables															
Age Left School	150,791	16.9	2.2	5	35	71,384	16.9	2.1	5	35	9,715	15	0	15	15
College	230,862	0.35	0.48	0	1	110,185	0.35	0.48	0	1	9,715	0	0	0	0
CSE	230,862	0.22	0.41	0	1	110,185	0.25	0.44	0	1	9,715	0.08	0.27	0	1
O-levels	230,862	0.54	0.50	0	1	110,185	0.55	0.50	0	1	9,715	0.12	0.33	0	1
Health Variables															
BMI	230,862	27.35	4.93	12.12	74.68	110,185	27.37	5.00	12.12	68.95	9,715	28.61	5.39	15.11	61.54
Diabetes	230,862	0.034	0.182	0	1	110,185	0.032	0.175	0	1	9,715	0.054	0.226	0	1
Control Variables															
Male	230,862	0.45	0.50	0.00	1.00	110,185	0.44	0.50	0.00	1.00	9,715	0.43	0.50	0.00	1.00
North Coord.	230,862	0	1	-2.39	2.44	110,185	0.00	1.00	-2.39	2.44	9,715	0.23	0.96	-2.24	2.44
East Coord.	230,862	0	1	-4.74	3.25	110,185	0.00	1.00	-4.74	3.25	9,715	-0.14	0.83	-3.96	3.25
Genetic Variables															
BMI Score	58,300	0	1	-4.08	3.89	27,902	0.00	1.00	-4.08	3.85	2,615	0.12	1.02	-3.87	3.66
Education Score	58,300	0	1	-3.84	3.92	27,902	0.00	1.00	-3.80	3.92	2,615	-0.30	0.96	-3.34	2.74
Genotyped	230,862	0.25	0.43	0	1	110,185	0.25	0.43	0	1	9,715	0.27	0.44	0	1

Note: This table reports the summary statistics for the variables used in this paper. Panels A and B report these values for a sample of individuals born within 10 years and 5 years of the discontinuity date, respectively. Panel C gives the same statistics for those born in the 5 years before the discontinuity date who report having dropped out of school at age 15. The variable *Age Left School* is only available in the subsample who do not report having a college degree. North Coordinate, East Coordinate, and both genetic score variables are measured in standard deviation units of the sample from Panel A. Genetic scores are only available in approximately one-quarter of the sample.

Table 2: McCrary Test and Balance Tests

	McCrary Test	Balance Tests			
	(1)	(2)	(3)	(4)	(5)
	Count	North Coord.	East Coord.	BMI Score	EA Score
After	-9.6	-0.016	0.009	0.026	0.017
	(19.2)	(0.014)	(0.011)	(0.022)	(0.021)
Linear Trend	X	X	X	X	X
Bandwidth	58	66	62	91	93
N	117	121,663	113,843	42,751	43,735

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the results of a McCrary test and balance tests for predetermined variables. In the balance test for the BMI and educational achievement scores, the data are restricted to those for whom genetic data are available. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Standard errors are clustered by month of birth.

Table 3: First-Stage Estimates

	(1) Age 14	(2) Age 14	(3) Age 15	(4) Age 15	(5) Age 16	(6) Age 16	(7) Age 17	(8) Age 17	(9) Age 18	(10) Age 18
After	-1.56e-05	9.54e-06	0.00855**	0.00625***	0.151***	0.139***	0.0158	0.00228	0.00848	0.00278
	(0.000276)	(0.000270)	(0.00400)	(0.00228)	(0.0100)	(0.00515)	(0.00981)	(0.00555)	(0.00793)	(0.00435)
Linear Trend	X	X	X	X	X	X	X	X	X	X
Controls		X		X		X		X		X
Fraction in School	99.9%	99.9%	99.4%	99.4%	90.0%	90.0%	61.9%	61.9%	53.5%	53.5%
Bandwidth	127	127	63	63	50	50	50	50	51	51
N	246,898	246,898	115,746	115,746	91,451	91,451	91,451	91,451	93,314	93,314
R2	0.000	0.001	0.002	0.012	0.056	0.079	0.000	0.062	0.000	0.062

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the first stage estimate of the effect of the 1972 ROSLA on the fraction of students still in school at various ages. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Table 4: Mean Effect of Education on BMI

	OLS		Reduced-Form		2SLS	
BMI	(1)	(2)	(3)	(4)	(5)	(6)
Edu16	-1.263***	-1.081***			-0.421	-0.525
	(0.0475)	(0.0471)			(0.367)	(0.371)
After			-0.0548	-0.0646		
			(0.0472)	(0.0453)		
Linear Trend	X	X	X	X	X	X
Controls		X		X		X
N	208,129	208,129	208,129	208,129	208,129	208,129
R2	0.008	0.044	0.001	0.040	0.005	0.043

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports OLS, reduced-form, and 2SLS estimates of the effect of the 1972 ROSLA on BMI. The indicator variable *Edu16* identifies individuals still in school at age 16. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012) (110 months). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Table 5: Mean Effect of Education on Diabetes

	OLS		Reduced-Form		2SLS	
Diabetes	(1)	(2)	(3)	(4)	(5)	(6)
Edu16	-0.019***	-0.018***			-0.023*	-0.030***
	(0.0019)	(0.0019)			(0.012)	(0.011)
After			-0.00302*	-0.0037***		
			(0.0015)	(0.0014)		
Linear Trend	X	X	X	X	X	X
Controls		X		X		X
N	210,302	210,302	210,302	210,302	210,302	210,302
R2	0.004	0.011	0.002	0.010	0.004	0.010

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports OLS, reduced-form, and 2SLS estimates of the effect of the 1972 ROSLA on diabetes. The indicator variable *Edu16* identifies individuals still in school at age 16. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012) (111 months). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Table 6: Genetic Heterogeneity in the First-Stage

Edu16	(1)	(2)	(3)	(4)	(5)	(6)
After	0.160***	0.148***	0.148***			
	(0.0111)	(0.00600)	(0.00605)			
BMI Score x After	0.0245***	0.0235***	0.0230***			
	(0.00473)	(0.00458)	(0.00460)			
Low BMI Score x After				0.134***	0.124***	0.125***
				(0.0132)	(0.00972)	(0.00973)
Mid BMI Score x After				0.160***	0.148***	0.147***
				(0.0120)	(0.00910)	(0.00892)
High BMI Score x After				0.185***	0.174***	0.173***
				(0.0143)	(0.00866)	(0.00870)
Linear Trend	X	X	X	X	X	X
Controls		X	X		X	X
Geography x After			X			X
N	86,036	86,036	86,036	86,036	86,036	86,036
R2	0.058	0.080	0.088	0.058	0.080	0.088

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the heterogeneity of the first stage estimates of the effect of the 1972 ROSLA on the fraction of students still in school at age 16 (*Edu16*). In the categorical specification, *Low*, *Mid*, and *High BMI Score* correspond to binary variables identifying which tercile of the BMI score distribution the individual is in. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. In the categorical specification, the categorical variable is also included directly as a control. Standard errors are clustered by month of birth.

Table 7: Genetic Heterogeneity in the Reduced-Form for BMI

BMI	(1)	(2)	(3)	(4)	(5)	(6)
After	-0.0714	-0.0757	-0.0788			
	(0.0664)	(0.0642)	(0.0643)			
BMI Score x After	0.0661	0.0633	0.0626			
	(0.0489)	(0.0480)	(0.0478)			
Low BMI Score x After				-0.197**	-0.201**	-0.207**
				(0.0891)	(0.0854)	(0.0861)
Mid BMI Score x After				0.0199	0.0242	0.0252
				(0.0962)	(0.0916)	(0.0912)
High BMI Score x After				-0.0192	-0.0328	-0.0376
				(0.0952)	(0.0931)	(0.0926)
Linear Trend	X	X	X	X	X	X
Controls		X	X		X	X
Geography x After			X			X
Mean BMI (Low Score)				26.1	26.1	26.1
Mean BMI (Mid Score)				27.4	27.4	27.4
Mean BMI (High Score)				28.9	28.9	28.9
N	208,129	208,129	208,129	208,129	208,129	208,129
R2	0.020	0.040	0.041	0.016	0.036	0.037

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the heterogeneity of the reduced-form estimates of the effect of the 1972 ROSLA on mean BMI. In the categorical specification, *Low*, *Mid*, and *High BMI Score* correspond to binary variables identifying which tercile of the BMI score distribution the individual is in. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. In the categorical specification, the categorical variable is also included directly as a control. Standard errors are clustered by month of birth.

Table 8: Genetic Heterogeneity in the Reduced-Form for the 90th Percentile of BMI

90th Percentile BMI	(1)	(2)	(3)	(4)	(5)	(6)
After	-0.161	-0.208*	-0.157			
	(0.152)	(0.121)	(0.124)			
BMI Score x After	0.257**	0.223**	0.224***			
	(0.110)	(0.0884)	(0.0826)			
Low BMI Score x After				-0.489**	-0.388***	-0.460***
				(0.200)	(0.141)	(0.138)
Mid BMI Score x After				-0.211	-0.213	-0.188
				(0.224)	(0.224)	(0.192)
High BMI Score x After				0.149	0.148	0.115
				(0.246)	(0.225)	(0.198)
Linear Trend	X	X	X	X	X	X
Controls		X	X		X	X
Geography x After			X			X
90th ptile BMI (Low Score)				31.6	31.6	31.6
90th ptile BMI (Mid Score)				33.6	33.6	33.6
90th ptile BMI (High Score)				36.0	36.0	36.0
N	208,129	208,129	208,129	208,129	208,129	208,129

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the heterogeneity of the reduced-form estimates of the effect of the 1972 ROSLA on the 90th percentile of BMI. In the categorical specification, *Low*, *Mid*, and *High BMI Score* correspond to binary variables identifying which tercile of the BMI score distribution the individual is in. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. In the categorical specification, the categorical variable is also included directly as a control. Standard errors are clustered by month of birth.

Table 9: Genetic Heterogeneity in the Reduced-Form for Diabetes

Diabetes	(1)	(2)	(3)	(4)	(5)	(6)
After	-0.0056**	-0.0062***	-0.0061***			
	(0.0022)	(0.0021)	(0.0021)			
BMI Score x After	-0.0053***	-0.0053***	-0.0052***			
	(0.0018)	(0.0017)	(0.0017)			
Low BMI Score x After				-0.00022	-0.00077	-0.00086
				(0.0029)	(0.0029)	(0.0029)
Mid BMI Score x After				-0.0064**	-0.0067**	-0.0067**
				-0.0032	(0.0031)	(0.0030)
High BMI Score x After				-0.010**	-0.011***	-0.011***
				(0.0040)	(0.0039)	(0.0039)
Linear Trend	X	X	X	X	X	X
Controls		X	X		X	X
Geography x After			X			X
Mean Diabetes (Low Score)				0.025	0.025	0.025
Mean Diabetes (Mid Score)				0.035	0.035	0.035
Mean Diabetes (High Score)				0.048	0.048	0.048
N	210,302	210,302	210,302	210,302	210,302	210,302
R2	0.004	0.010	0.011	0.004	0.009	0.010

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the heterogeneity of the reduced-form estimates of the effect of the 1972 ROSLA on diabetes. In the categorical specification, *Low*, *Mid*, and *High BMI Score* correspond to binary variables identifying which tercile of the BMI score distribution the individual is in. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. In the categorical specification, the categorical variable is also included directly as a control. Standard errors are clustered by month of birth.

Table 10: Genetic Heterogeneity in the 2SLS Estimates for BMI and Diabetes

	(1) BMI	(2) BMI	(3) BMI	(4) BMI	(5) Diabetes	(6) Diabetes	(7) Diabetes	(8) Diabetes
Edu16	-0.561	-0.632			-0.0338**	-0.0396***		
	(0.457)	(0.461)			(0.0150)	(0.0149)		
BMI Score x Edu16	0.466	0.457			-0.0284***	-0.0276***		
	(0.304)	(0.296)			(0.0105)	(0.0104)		
Low BMI Score x Edu16			-1.530**	-1.599**			-0.00442	-0.0113
			(0.676)	(0.667)			(0.0225)	(0.0226)
Mid BMI Score x Edu16			0.0729	0.0469			-0.0425**	-0.0467**
			(0.626)	(0.609)			(0.0207)	(0.0197)
High BMI Score x Edu16			-0.144	-0.268			-0.0572***	-0.0632***
			(0.540)	(0.538)			(0.0221)	(0.0221)
Linear Trend	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	208,129	208,129	208,129	208,129	210,302	210,302	210,302	210,302
R2	0.023	0.043	0.019	0.038	0.004	0.010	0.004	0.009

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the heterogeneity of the 2SLS estimates of the effect of remaining in school till age 16 (*Edu16*) on BMI and diabetes. In the categorical specifications, *Low*, *Mid*, and *High BMI Score* correspond to binary variables identifying which tercile of the BMI score distribution the individual is in. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. In the categorical specification, the categorical variable is also included directly as a control. Standard errors are clustered by month of birth.

Table 11: Selected Genetic Correlations with BMI

Trait	r_g	SE
Weight-related		
Childhood Obesity	0.732***	(0.046)
Extreme BMI	1.027***	(0.025)
Obesity Class 1	1.02***	(0.010)
Obesity Class 2	1.046***	(0.016)
Obesity Class 3	0.933***	(0.052)
Overweight	1.019***	(0.012)
Waist-Hip Ratio	0.579***	(0.025)
Diabetes-related		
Fasting Glucose	0.312***	(0.053)
Fasting Insulin	0.65***	(0.062)
Fasting Proinsulin	0.348	(0.163)
Type-2 Diabetes	0.361***	(0.044)
Education-related		
Childhood IQ	-0.173	(0.057)
College	-0.242***	(0.033)
Years of Education	-0.279***	(0.035)
Smoking-related		
Cigarettes per Day	0.287***	(0.073)
Ever/Never Smoked	0.204***	(0.041)
Former/Current Smoker	-0.185**	(0.054)
Height-related		
Extreme Height	-0.09	(0.033)
Height	-0.093***	(0.021)
Blood Results		
HDL	-0.382***	(0.039)
Total Cholesterol	0.023	(0.032)
Triglycerides	0.267***	(0.044)
LDL	0.082	(0.038)
Mental Health		
Bipolar	-0.071	(0.041)
Depression	-0.015	(0.064)
Schizophrenia	-0.095***	(0.025)
Other diseases		
Coronary Artery Disease	0.217***	(0.045)
Crohn's Disease	0.021	(0.038)
Rheumatoid Arthritis	-0.051	(0.046)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Notes: Genetic correlations (r_g) are drawn from Bulik-Sullivan et al (2015). Complete results and descriptions of each trait may be found in that paper. Significance values are based on Bonferroni-corrected p-values for the full set of traits tested in Bulik-Sullivan et al.

Table 12: The Common Contributions to Heterogeneity of the BMI and EA Genetic Scores

	(1) Edu16	(2) Edu16	(3) Edu16	(4) BMI	(5) BMI	(6) BMI	(7) Diabetes	(8) Diabetes	(9) Diabetes
After	0.148***	0.148***	0.148***	-0.0757	-0.0758	-0.0758	-0.00615***	-0.00615***	-0.00615***
	(0.00600)	(0.00600)	(0.00603)	(0.0642)	(0.0642)	(0.0642)	(0.00214)	(0.00214)	(0.00214)
BMI Score x After	0.0235***			0.0633			-0.00532***		
	(0.00458)			(0.0480)			(0.00174)		
BMI Score Res x After		0.0171***	0.0172***		0.0636	0.0636		-0.00432**	-0.00432**
		(0.00462)	(0.00463)		(0.0477)	(0.0477)		(0.00176)	(0.00176)
EA Score x After			-0.0474***			-0.00702			0.00756***
			(0.00465)			(0.0511)			(0.00173)
Linear Trend	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X
N	86,036	86,036	86,036	208,129	208,129	208,129	210,302	210,302	210,302
R2	0.080	0.080	0.081	0.040	0.040	0.040	0.010	0.010	0.010

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table compares the continuous heterogeneity specification drawn from Table XXX to a specification where the genetic score for BMI is replaced by the residuals of the regression of the BMI score on the educational attainment score (*EA Score*). This residualized variable is denoted "*BMI Score Res*." The first-stage results (*Edu16*) and reduced-form results for BMI and diabetes are all reported. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Table 13: Bandwidth Analysis

(A) First Stage	(1)	(2)	(3)	(4)	(5)
After	0.151***	0.144***	0.139***	0.136***	0.133***
	(0.00666)	(0.00576)	(0.00519)	(0.00501)	(0.00491)
Bandwidth	23	35	47	59	71
N	40,965	62,405	84,272	106,581	128,903
R2	0.086	0.081	0.079	0.079	0.079

(B) BMI	(1)	(2)	(3)	(4)	(5)
After	-0.0595	-0.0490	-0.0646	-0.0815**	-0.0960**
	(0.0599)	(0.0505)	(0.0453)	(0.0412)	(0.0390)
Bandwidth	56	83	110	137	164
N	101,086	151,445	205,907	265,719	307,130
R2	0.041	0.040	0.040	0.039	0.038

(C) Diabetes	(1)	(2)	(3)	(4)	(5)
After	-0.00446**	-0.00353**	-0.00365***	-0.00374***	-0.00354***
	(0.00185)	(0.00158)	(0.00138)	(0.00126)	(0.00120)
Bandwidth	57	84	111	138	165
N	103,023	153,298	208,129	267,453	308,544
R2	0.009	0.009	0.010	0.010	0.011

*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table reports the first-stage estimate and reduced-form estimates for BMI and diabetes for a variety of bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth selected by Imbens and Kalyanaraman (2012). In each panel, column (3) corresponds to the optimal bandwidth. These estimates come from a local linear framework with a triangle kernel. In all specifications, controls are included for a linear trend that is allowed to vary before and after the discontinuity, sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Table 14: Local Linear versus Global Polynomial

	Edu16		BMI		BMI (90th ptile)		Diabetes		BMI		Diabetes	
BMI	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(5)	(6)
After	0.14***	0.14***	-0.065	-0.045	-0.23*	-0.30**	-0.0037***	-0.0069***				
	(0.0052)	(0.0064)	(0.045)	(0.069)	(0.12)	(0.13)	(0.0014)	(0.0021)				
Edu16									-0.53	-0.32	-0.030***	-0.050***
									(0.37)	(0.50)	(0.011)	(0.015)
Local Linear	X		X		X		X		X		X	
Global Polynomial		X		X		X		X		X		X
Controls	X	X	X	X	X	X	X	X	X	X	X	X
OLS	X	X	X	X			X	X				
Quantile Reg.					X	X						
2SLS									X	X	X	X
N	86,036	378,945	208,129	378,945	208,129	378,945	210,302	378,945	208,129	378,945	210,302	378,945
R2	0.079	0.146	0.040	0.035			0.010	0.017	0.043	0.038	0.010	0.014

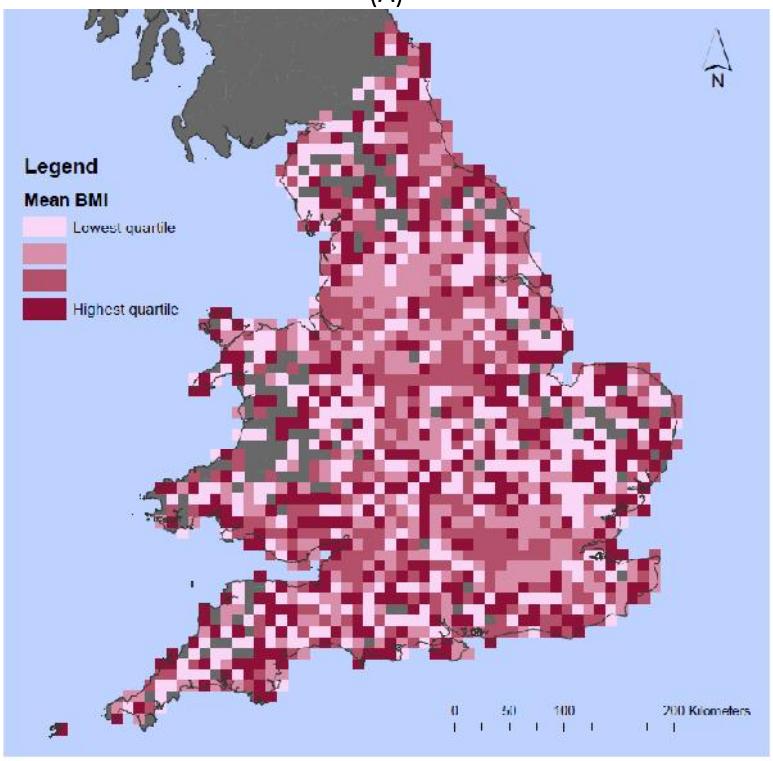
*** p < 0.01, ** p < 0.05, * p < 0.10

Note: This table compares the local linear specification used as the preferred specification in this paper to a global polynomial approach. In the local linear approach, trends are modeled as linear function that can vary before and after the threshold, and the data are weighted with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In the global polynomial approach, all data are included but the trends before and after the discontinuity are modeled as cubic polynomials. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

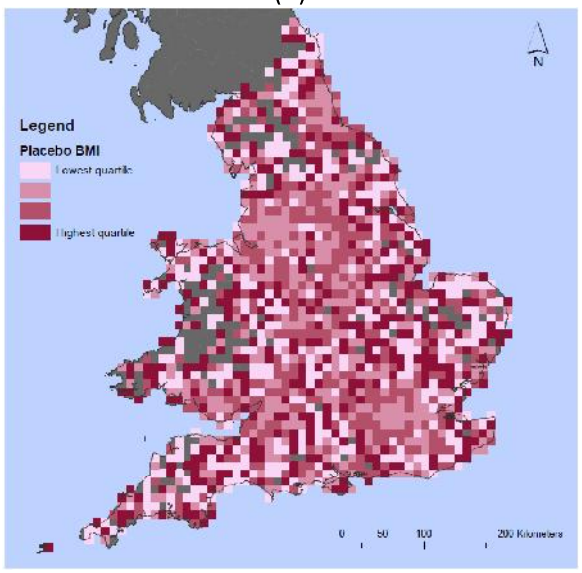
Figures

Figure 1: Geographic Distribution of the BMI Genetic Score

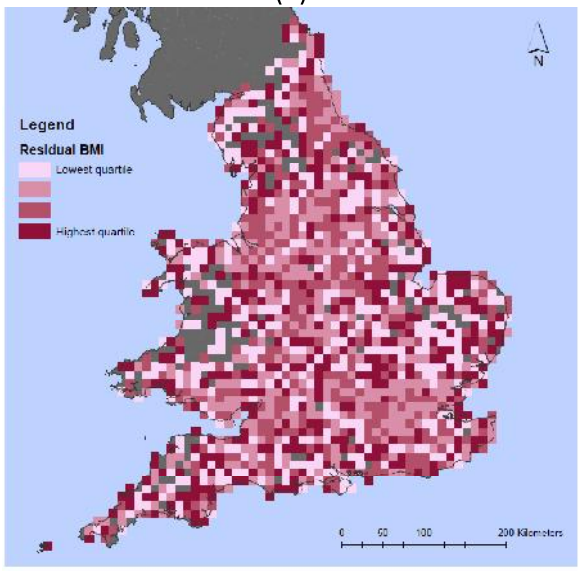
(A)



(B)

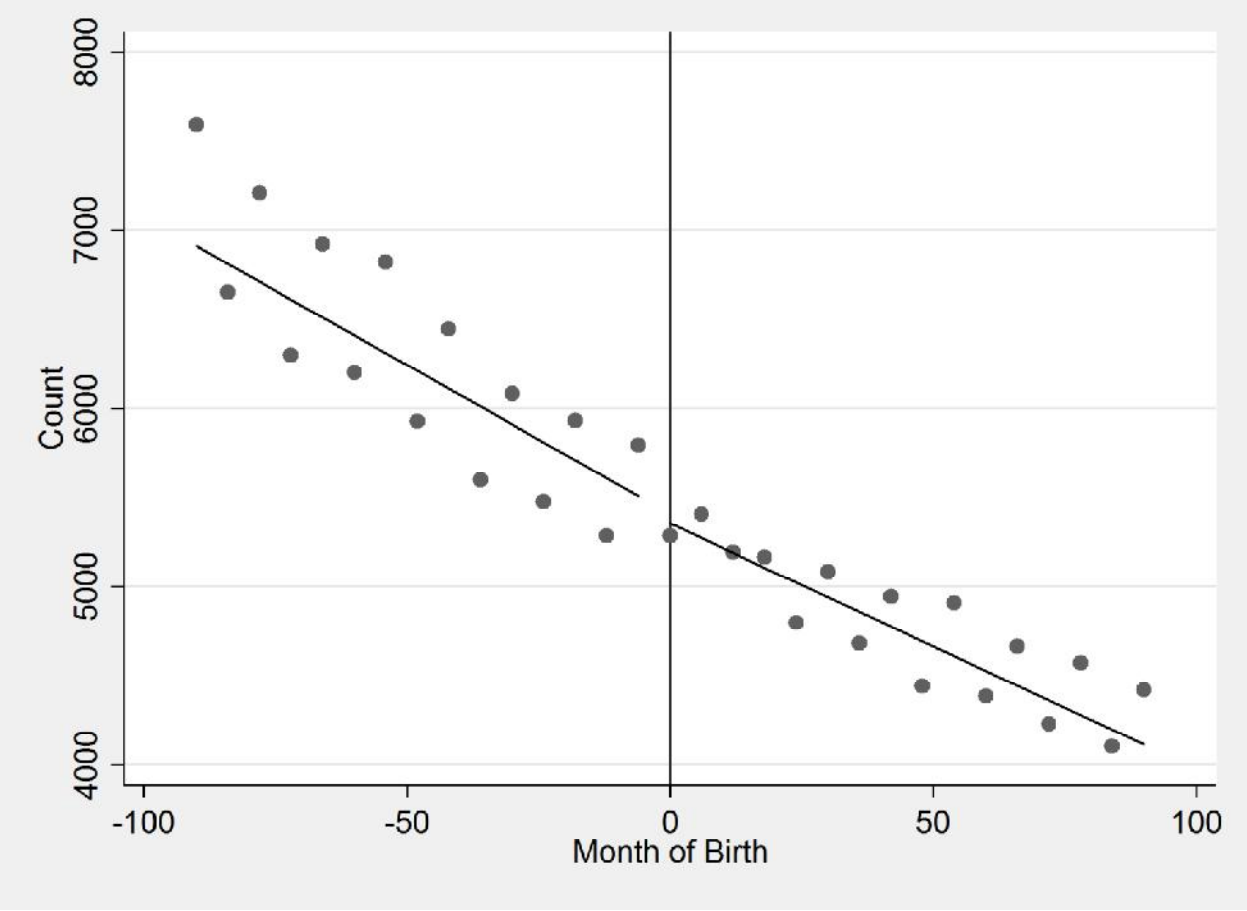


(C)



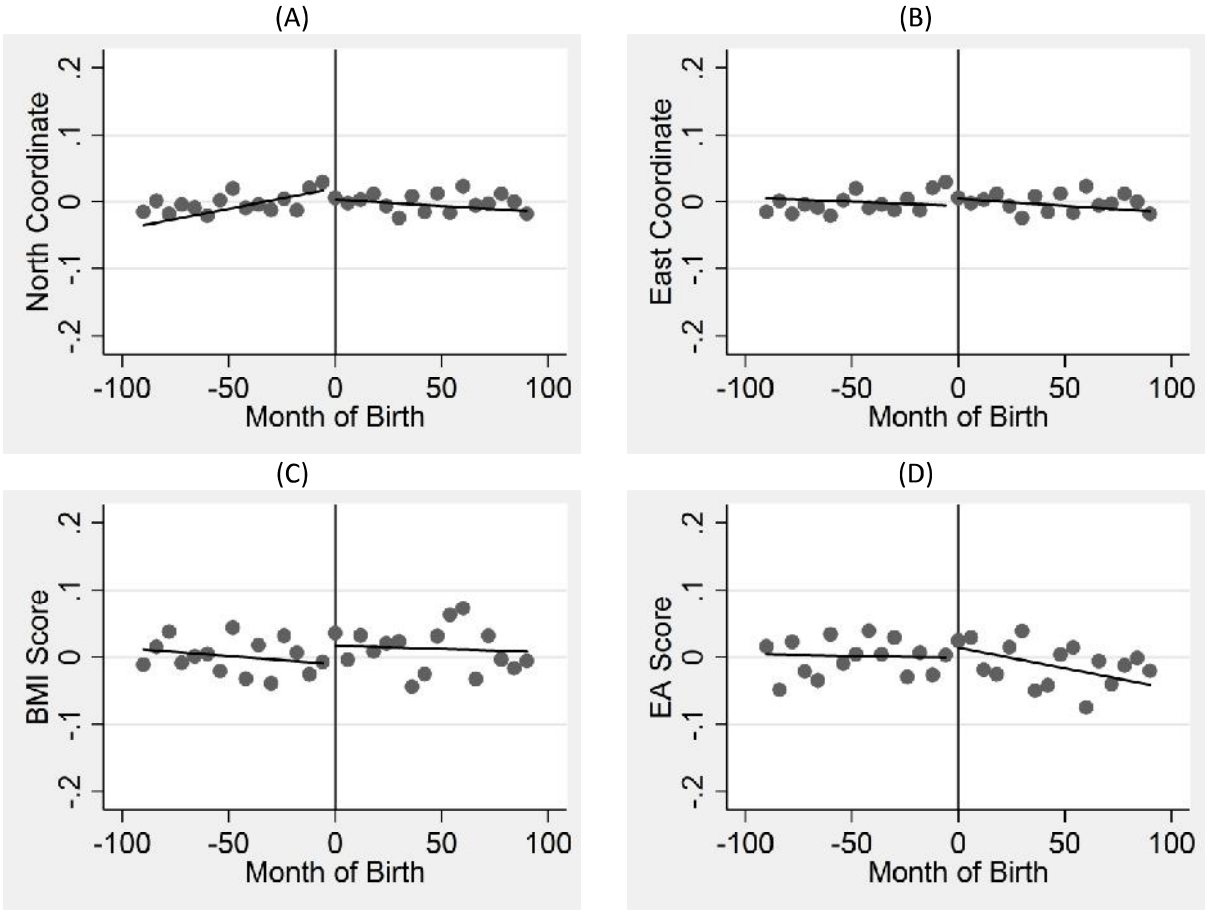
Note: These figures plot the geographic distribution of the genetic BMI score. For each square of the grid, the BMI score for all individuals born in that square is averaged and color coded according to which its quartile of average BMI score. Panel A performs this for analysis for the raw BMI scores. Panel B shows the resulting figure when we first randomly permute every individual's BMI score. Panel C shows the geographic distribution of the BMI score residual after controlling for county fixed effects and a quadratic of latitude and longitude.

Figure 2: McCrary Test Scatter Plot



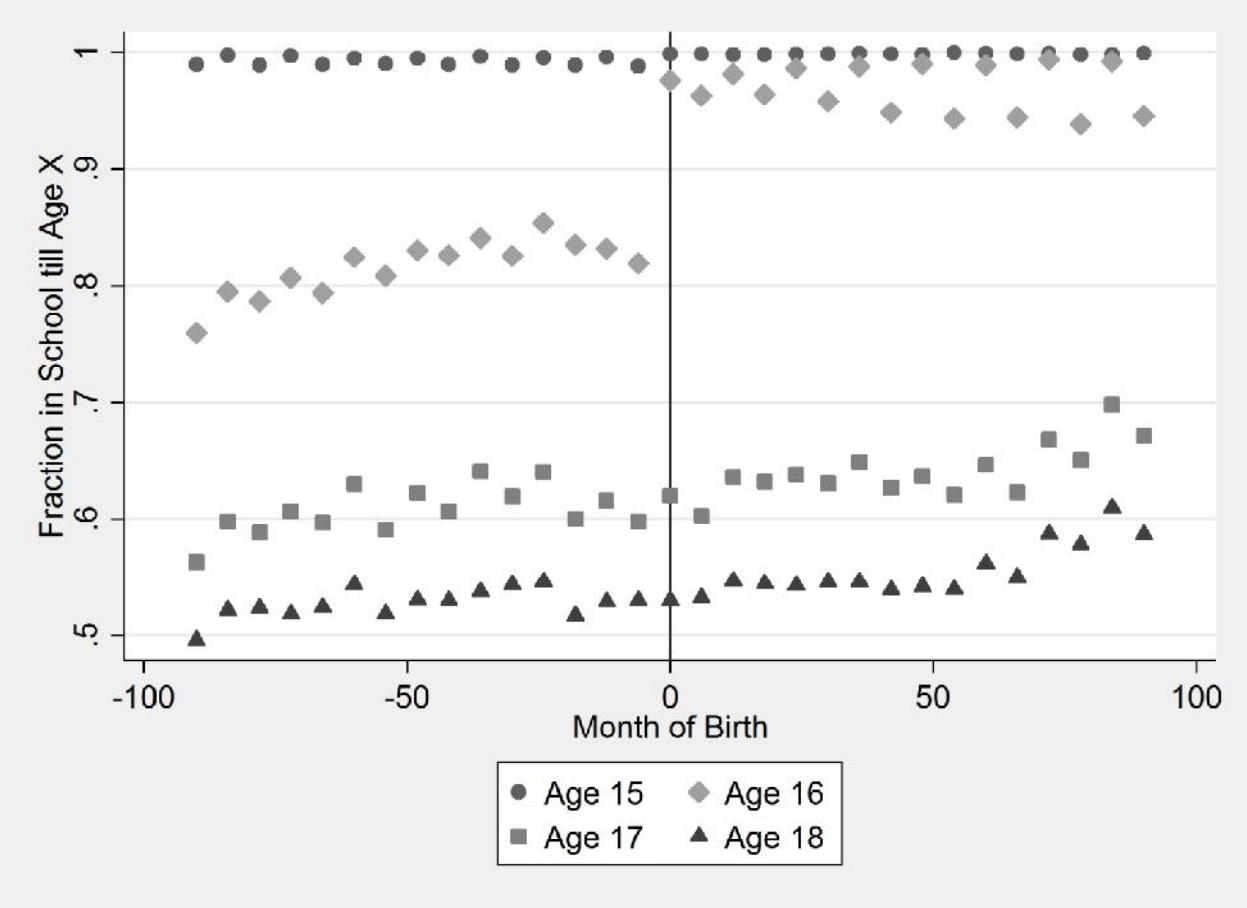
Note: This figure plots the distribution of *Month of Birth* in a window of the discontinuity. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who weren't. Each point corresponds to the number of individuals in our data set born within a 6-month bin.

Figure 3: Balance Tests



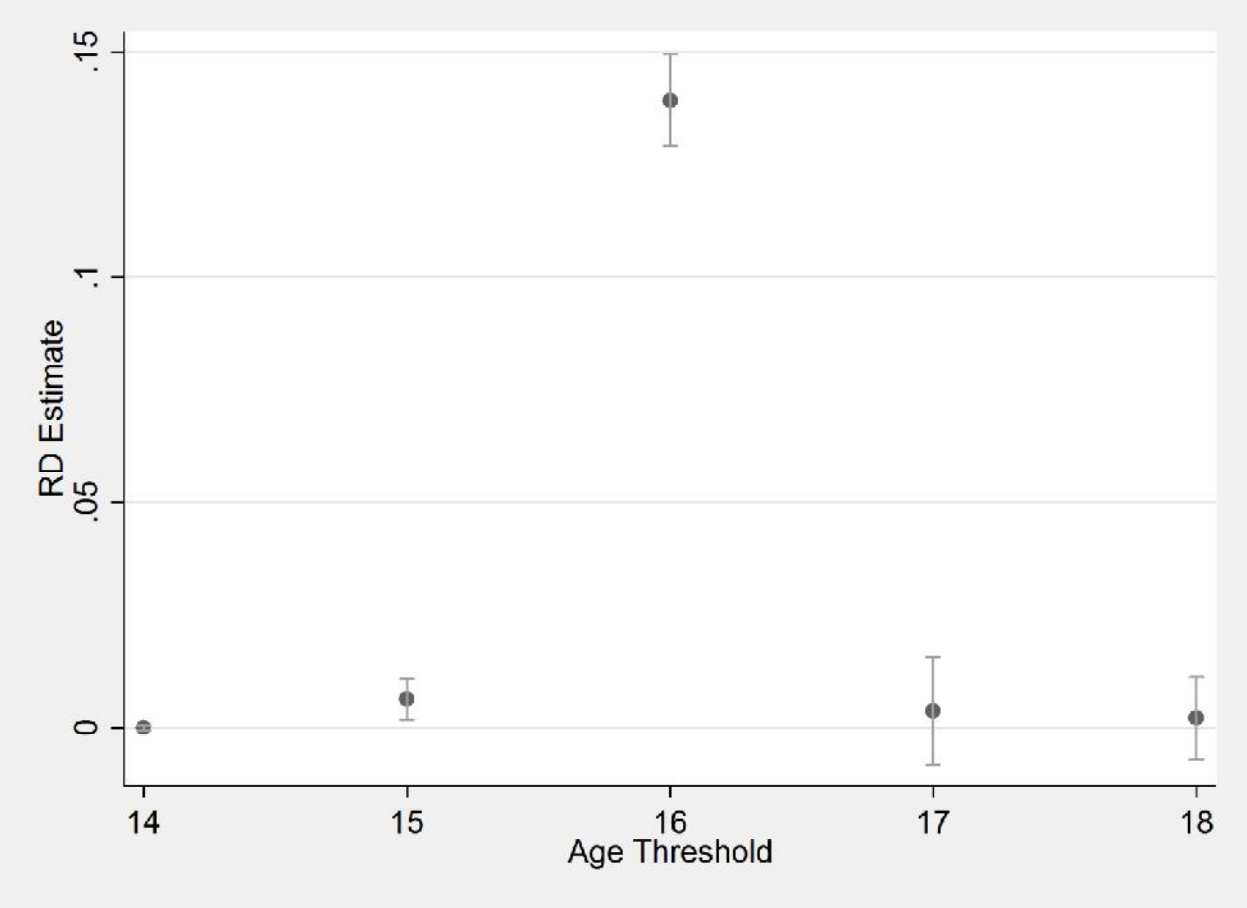
Notes: This figure contains scatter plots of the mean of four outcomes that were determined before the implementation of the ROSLA against date of birth. Panels A and B correspond to the North and East Coordinates of place of birth, measured in standard deviation units, and Panels C and D correspond to the BMI polygenic score (*BMI Score*) and educational attainment scores (*EA Score*), also measured in standard deviation units. For these scatter plots, individual outcomes are grouped into 6-month bins. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who were not.

Figure 4: Fraction in School till Various Ages by Date of Birth



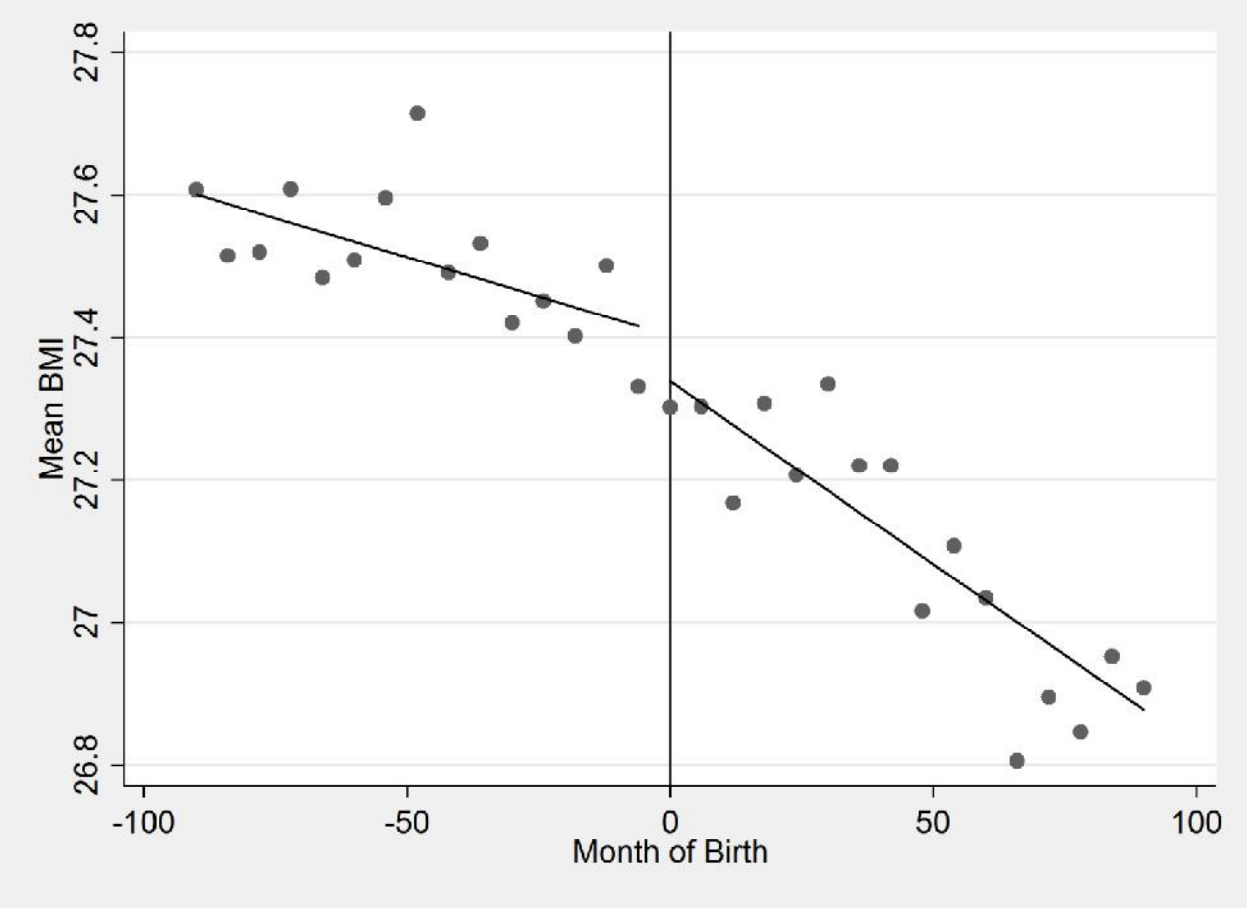
Notes: This figure shows a scatter plot of the fraction of individuals who remained in school till various ages against date of birth. More precisely, a student is coded as having stayed in school till age *X* if they report leaving school at age *X* or later. For this scatter plot, individual outcomes are grouped into 6-month bins. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who were not. Since respondents in our data only report the age they left school if they do not have a college or university degree, individuals with a college degree are assumed to have left school at age 18 or later.

Figure 5: RD Estimates for Fraction in School till Various Ages



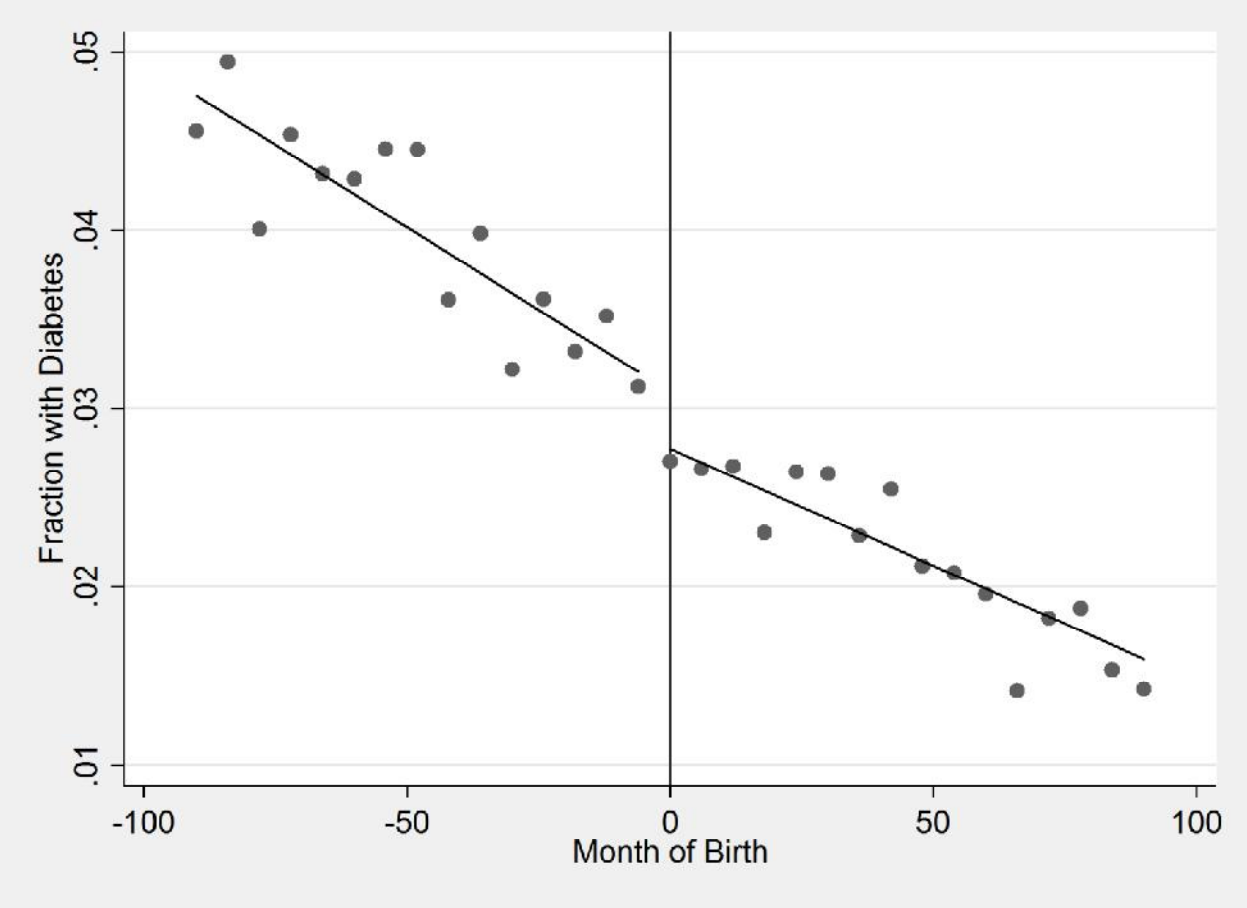
Notes: This figure plots regression discontinuity estimates and their 95% confidence intervals of the impact of the 1972 ROSLA on the fraction of individuals staying in school till various ages. For our binary outcome variable, a student is coded as having stayed in school till age *X* if they report leaving school at age *X* or later. Since respondents in our data only report the age they left school if they do not have a college or university degree, individuals with a college degree are assumed to have left school at age 18 or later. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 6: Mean BMI by Date of Birth



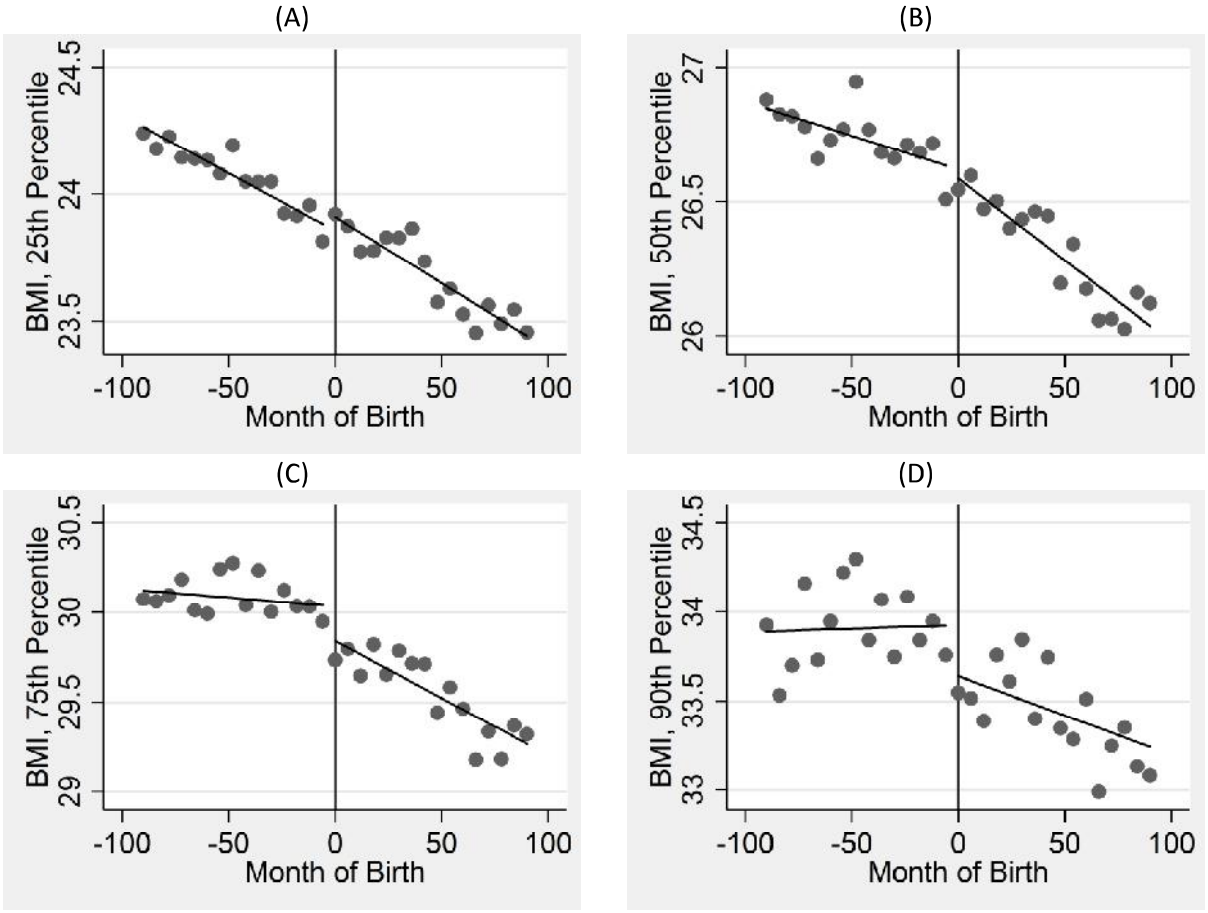
Notes: This figure shows a scatter plot of the mean BMI against date of birth. For this scatter plot, individual outcomes are grouped into 6-month bins. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who were not.

Figure 7: Fraction with Self-reported Diabetes by Date of Birth



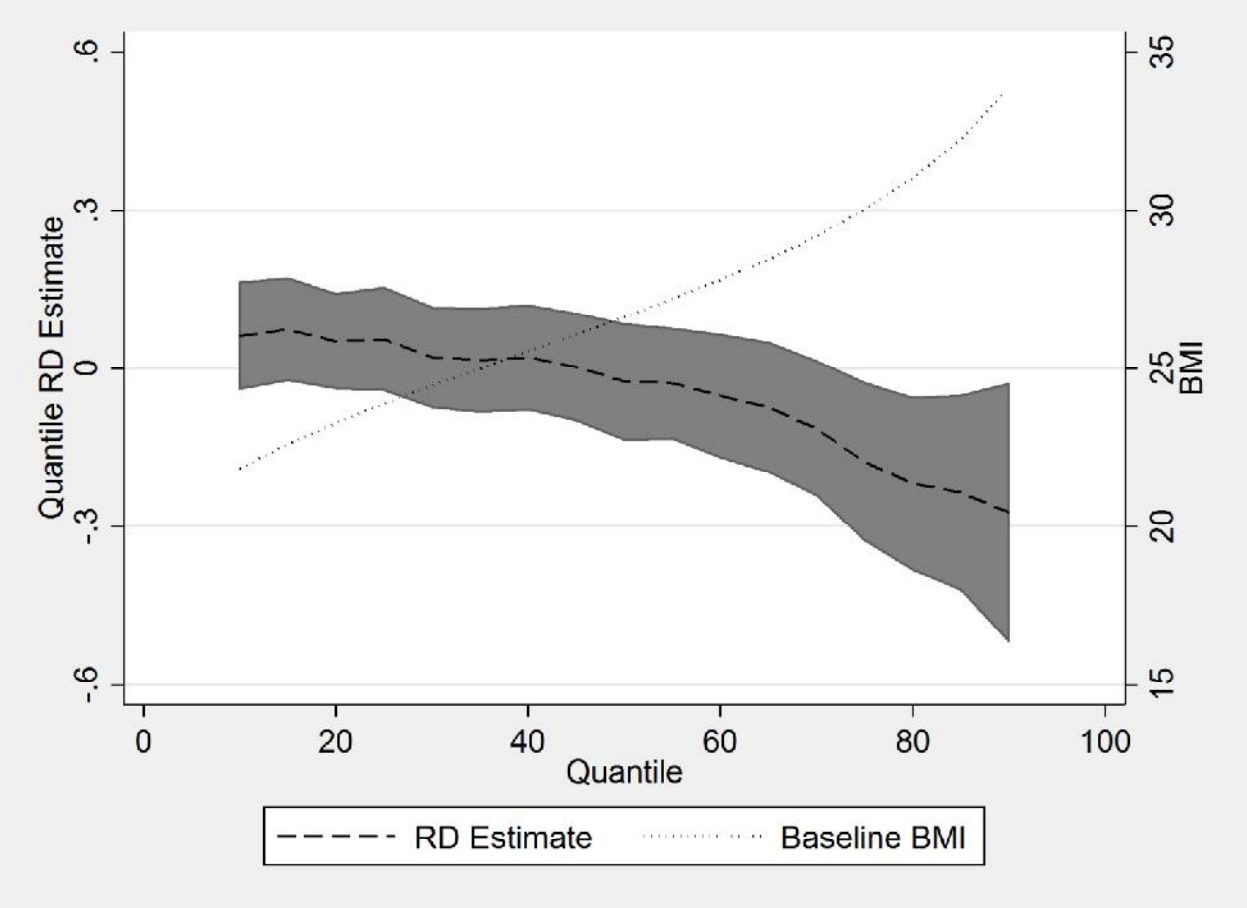
Notes: This figure shows a scatter plot of the fraction of individuals reporting that they have been diagnosed with diabetes against date of birth. For this scatter plot, individual outcomes are grouped into 6-month bins. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who were not.

Figure 8: Various Quantiles of BMI by Date of Birth



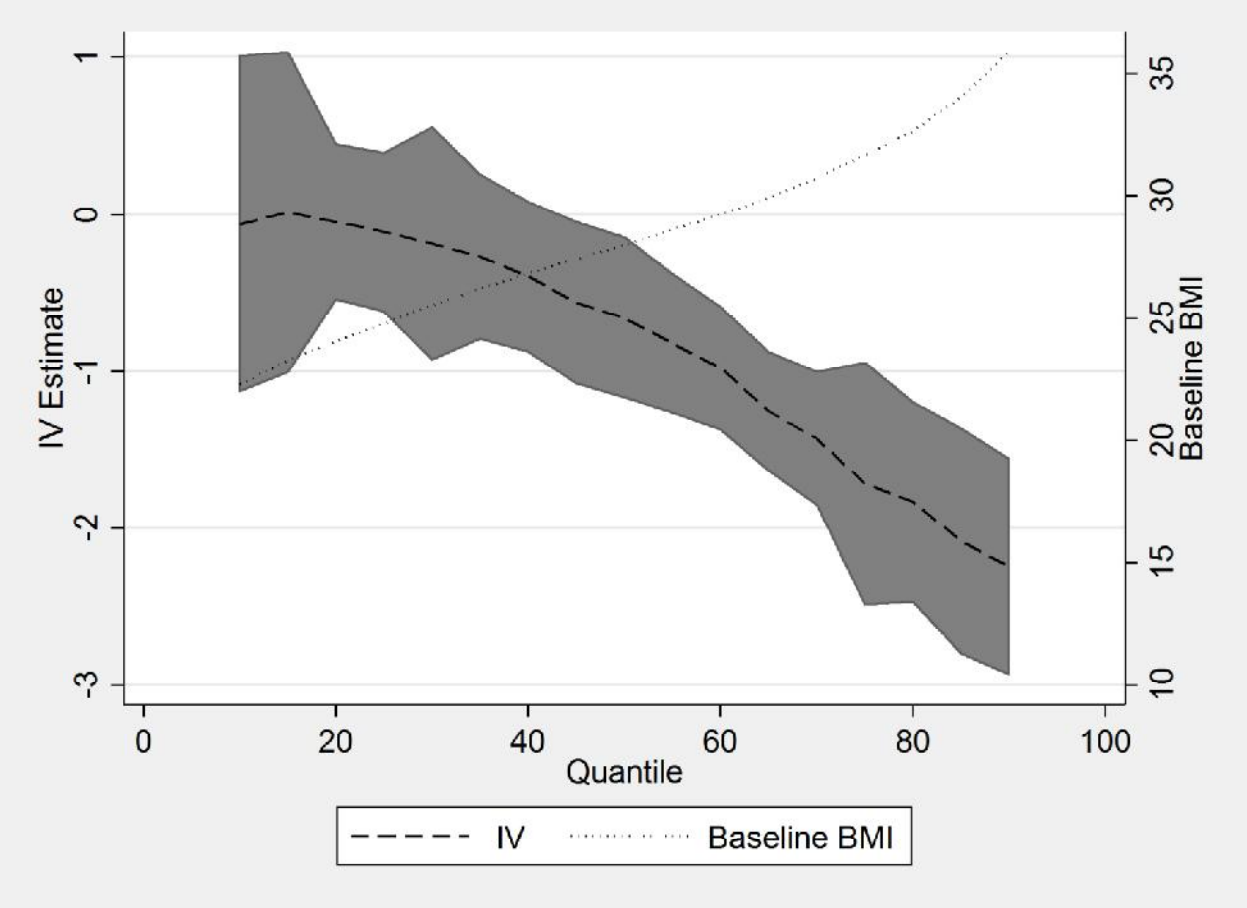
Notes: This figure shows a scatter plot of various percentiles of BMI against date of birth. For this scatter plot, individual outcomes are grouped into 6-month bins. The x-axis is measured in months since September 1957, the relevant threshold dividing those who were affected by the 1972 ROSLA and those who were not.

Figure 9: The Effect of the 1972 ROSLA on the Distribution of BMI



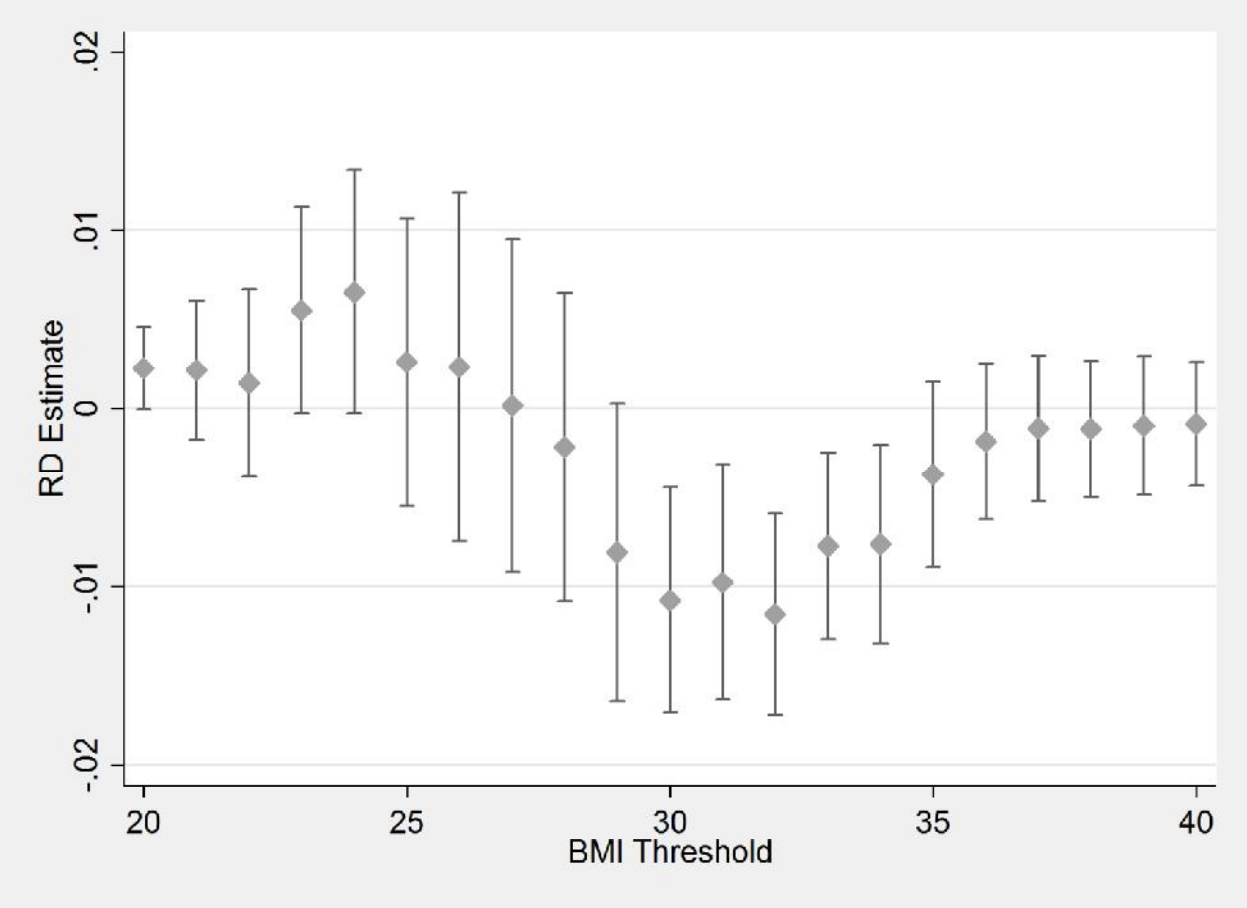
Notes: This figure displays estimates of the impact of the 1972 ROSLA on different quantiles of the BMI distribution in the whole population. The dashed line and gray band correspond to the estimates and 95% confidence intervals from a series of quantile regressions. The dotted line corresponds to the quantiles of BMI before the implementation of the ROSLA. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth.

Figure 10: The Effect of Staying in School Till Age 16 on the Distribution of BMI in Compliers



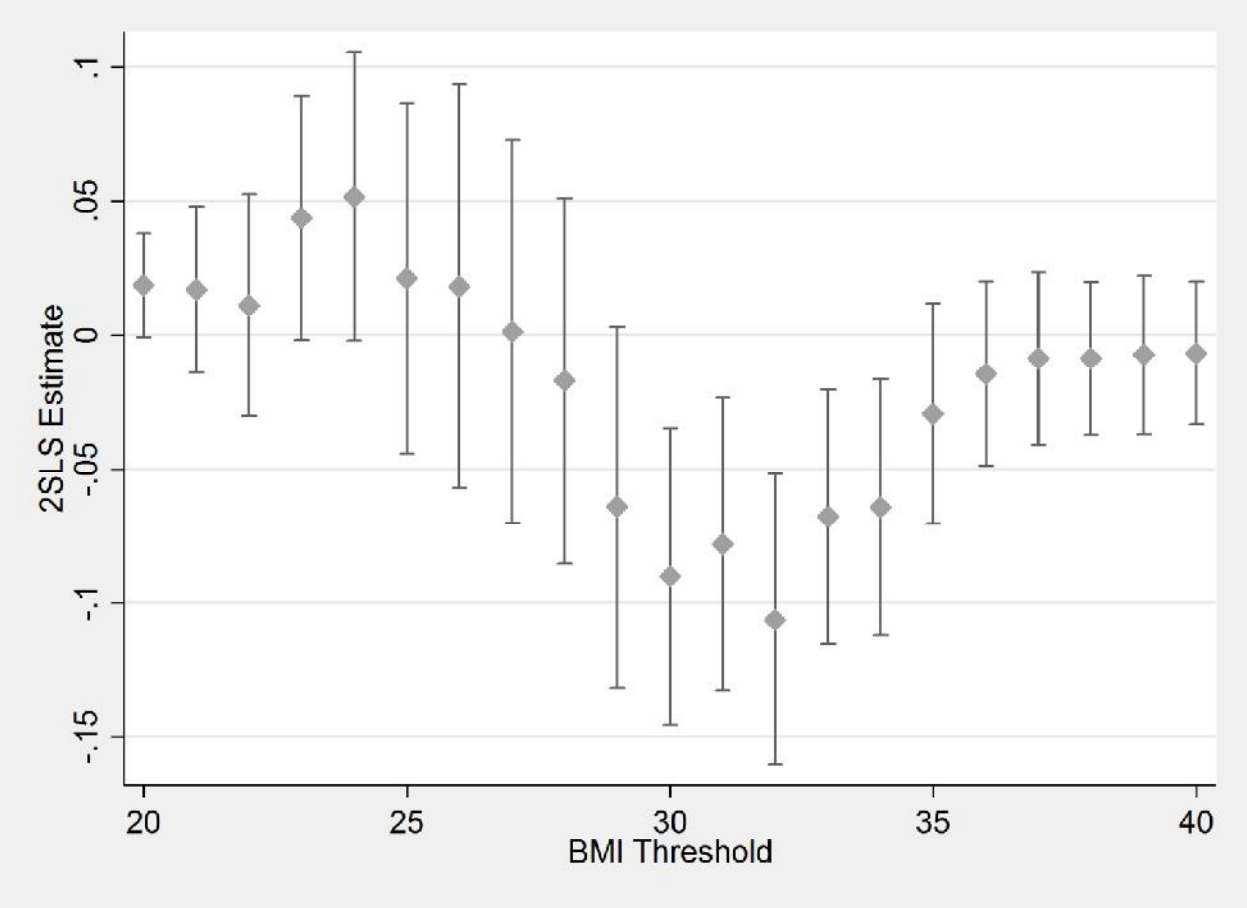
Notes: This figure displays estimates of the impact of staying in school till age 16 on various quantiles of BMI for the population of compliers (i.e. those who would have left school before their 16th birthday if they had not been compelled to stay by the 1972 ROSLA). The dashed line and gray band correspond to the estimates and 95% confidence intervals, estimated by the procedure in Frandsen, Froelich, and Melly (2008). The dotted line corresponds to the quantiles of BMI of compliers before the implementation of the ROSLA.

Figure 11: RD Estimates for the Fraction of Individuals with a BMI above Various Thresholds



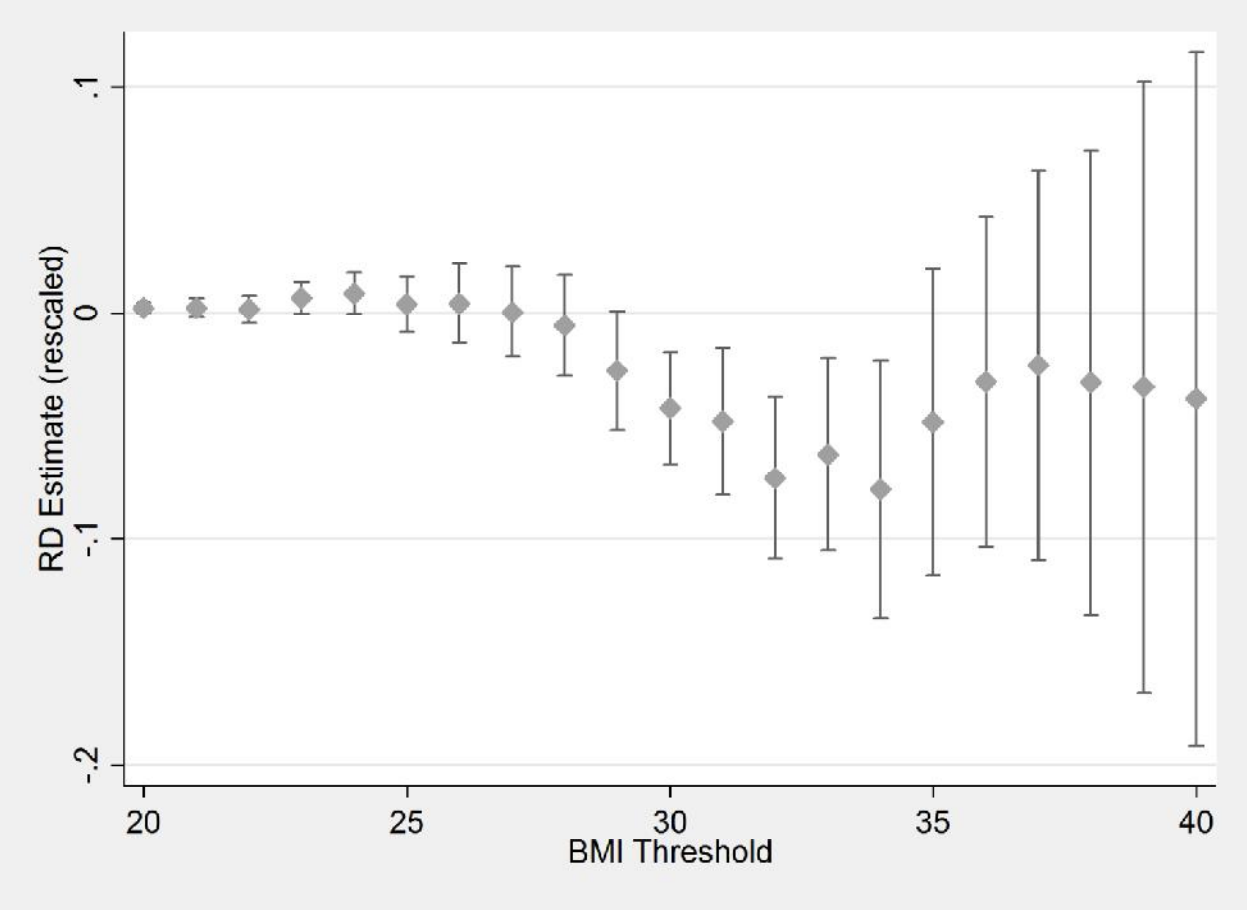
Notes: This figure displays estimates of the reduced-form impact of the 1972 ROSLA on the fraction of individuals with a BMI above various thresholds. Each point estimate and 95% confidence interval correspond to a separate RD regression on an indicator variable of whether an individual has a BMI greater than or equal to the relevant threshold. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 12: 2SLS Estimates for the Fraction of Individuals with a BMI above Various Thresholds



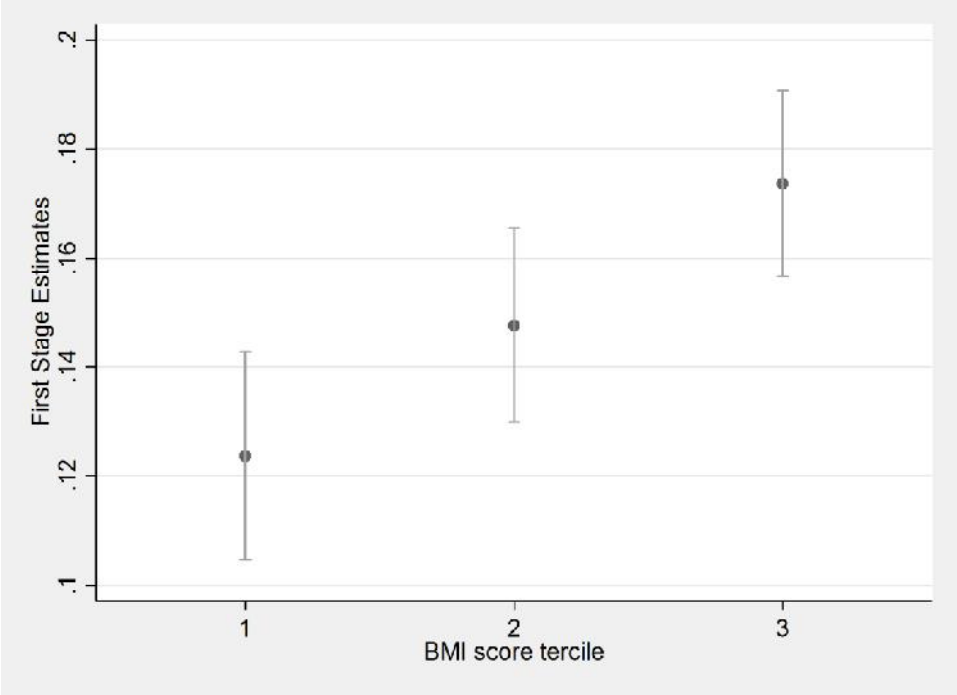
Notes: This figure contains the 2SLS estimates of the impact of staying in school till age 16 on having a BMI above various thresholds. Each point estimate and 95% confidence interval correspond to a separate 2SLS RD regression. The outcome variable for each regression is an indicator of whether an individual has a BMI greater than or equal to the relevant threshold. The endogenous regressor of whether the individual was in school till at least age 16 is instrumented by whether the individual was born after September 1, 1957 and would therefore be potentially constrained by the 1972 ROSLA. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). In all specifications, the linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 13: RD Estimates for the Fraction of Individuals with a BMI above Various Thresholds (Rescaled)



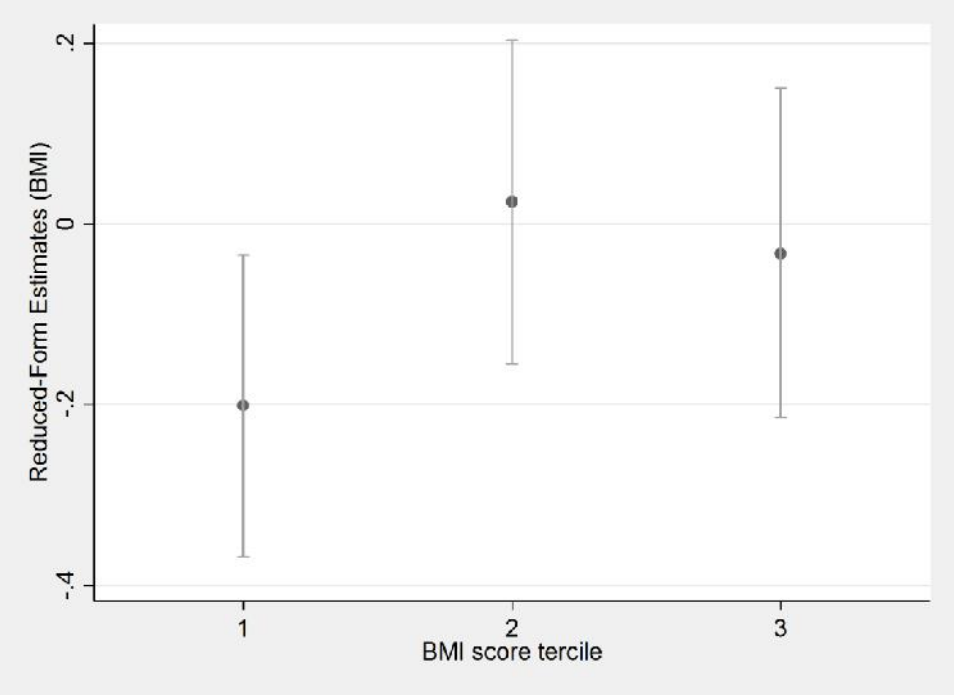
Notes: This figure plots the estimates from Figure 12 after dividing the estimates by the fraction of individuals in the sample with a BMI above the associated threshold. This changes the interpretation of these estimates from the *percentage-point decrease* in the fraction of individuals with a BMI above the threshold to the *percentage decrease* in the fraction of individuals with a BMI above the threshold.

Figure 14: First-Stage Estimates by Genetic BMI Score Tercile



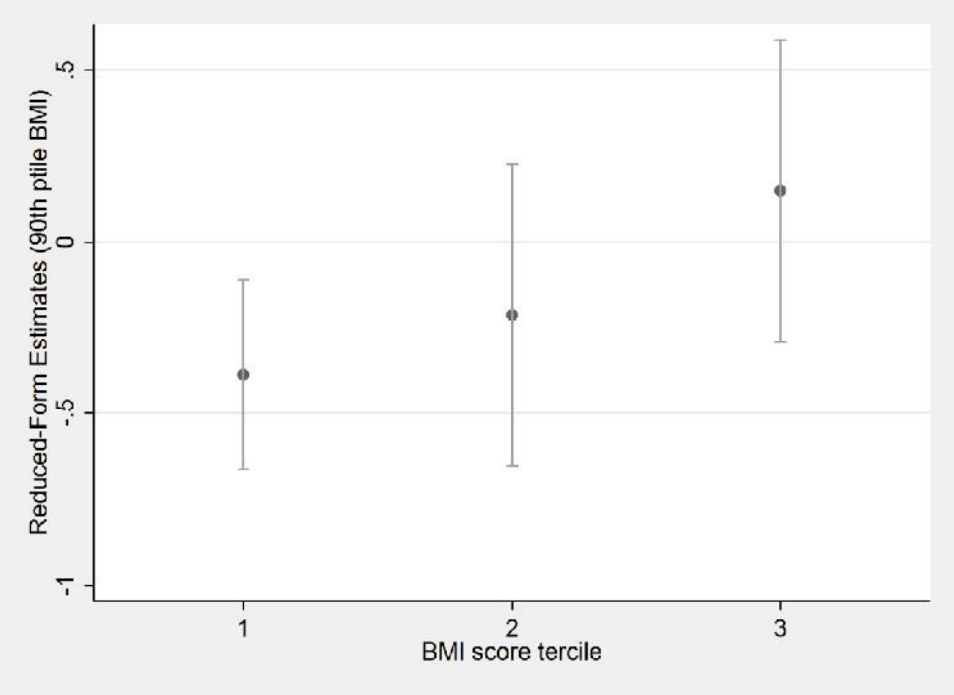
Notes: This figure reports the reduced-form estimates of the heterogeneous impact of the 1972 ROSLA on the fraction of individuals remaining in school till at least age 16, separated by the tercile of the individual's BMI score. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 15: Reduced-Form Estimates for BMI by Genetic BMI Score Tercile



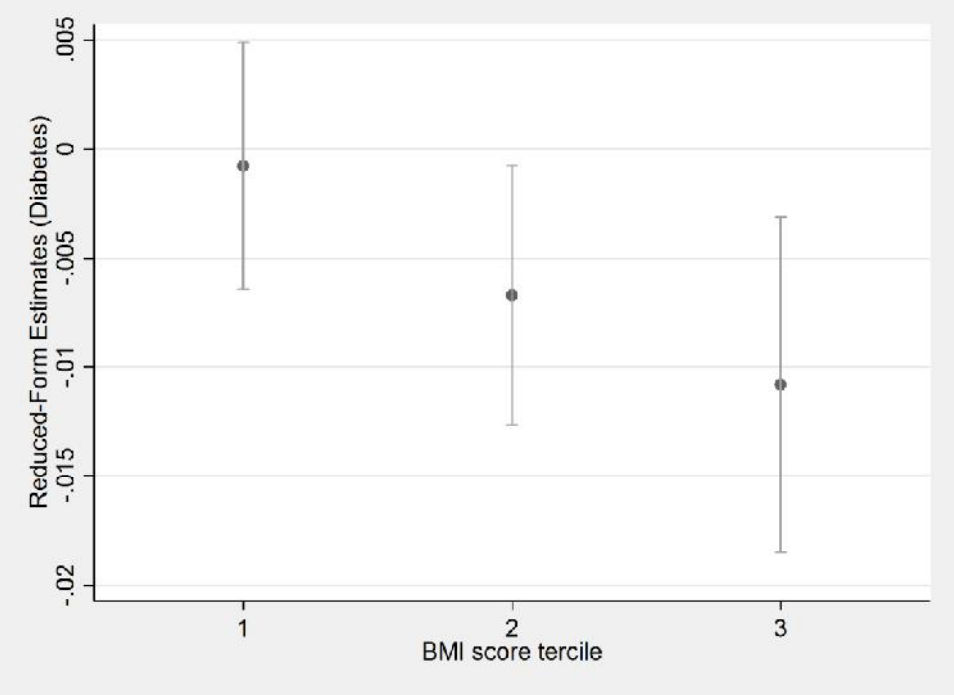
Notes: This figure reports the reduced-form estimates of the heterogeneous impact of the 1972 ROSLA on mean BMI, separated by the tercile of the individual's BMI score. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 16: Reduced-Form Estimates for the 90th Percentile of BMI by Genetic BMI Score Tercile



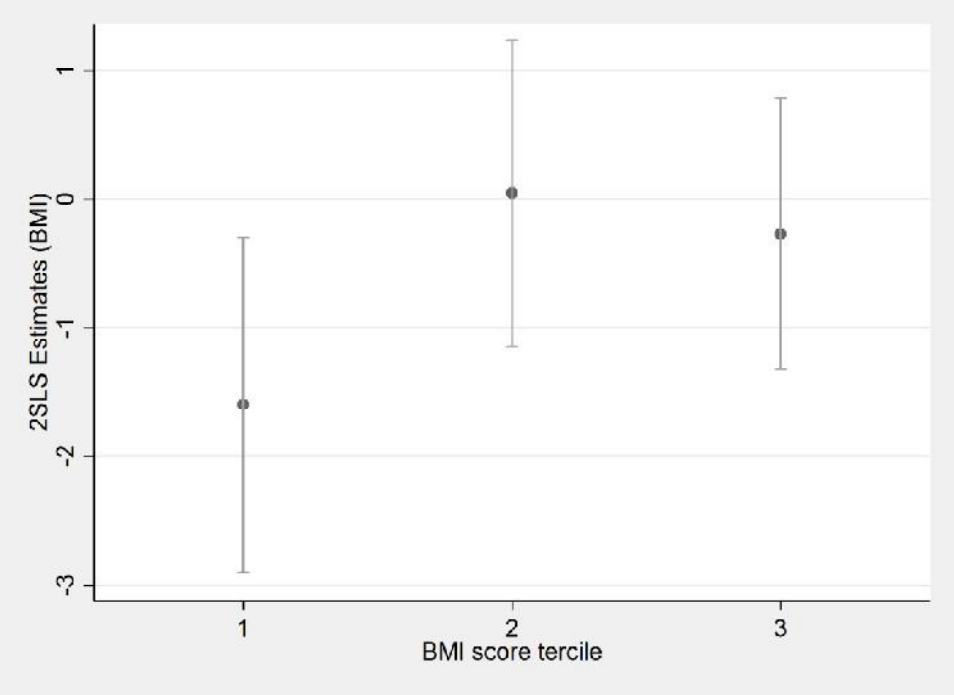
Notes: This figure reports the reduced-form estimates of the heterogeneous impact of the 1972 ROSLA on the 90th percentile of BMI, separated by the tercile of the individual's BMI score. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 17: Reduced-Form Estimates for Diabetes by Genetic BMI Score Tercile



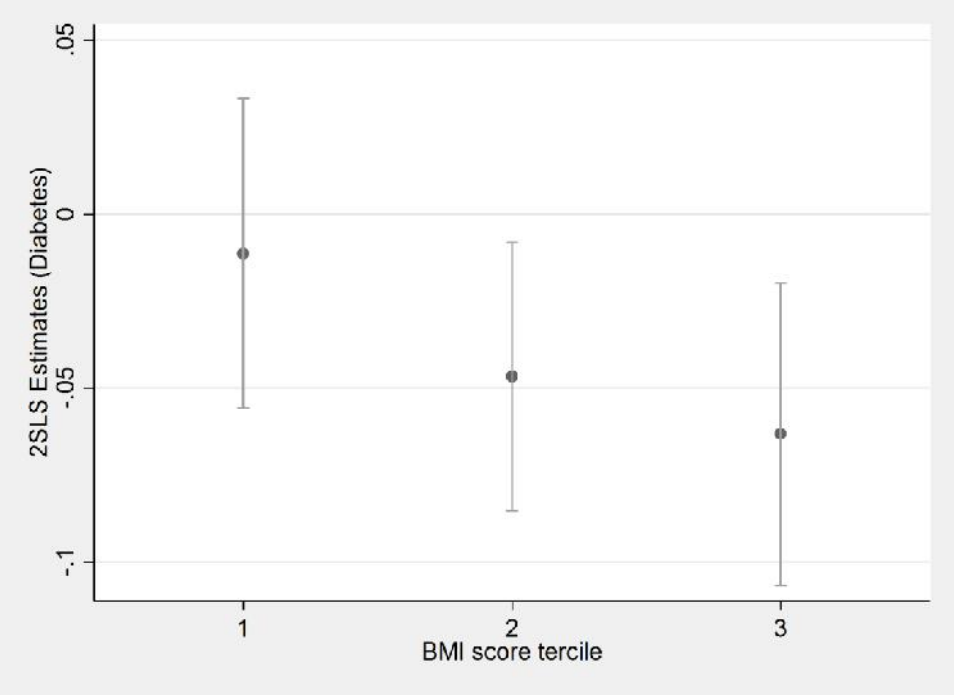
Notes: This figure reports the reduced-form estimates of the heterogeneous impact of the 1972 ROSLA on the fraction of individuals with diabetes, separated by the tercile of the individual's BMI score. These estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 18: The Effect of Staying in School till Age 16 on BMI by BMI Score Tercile



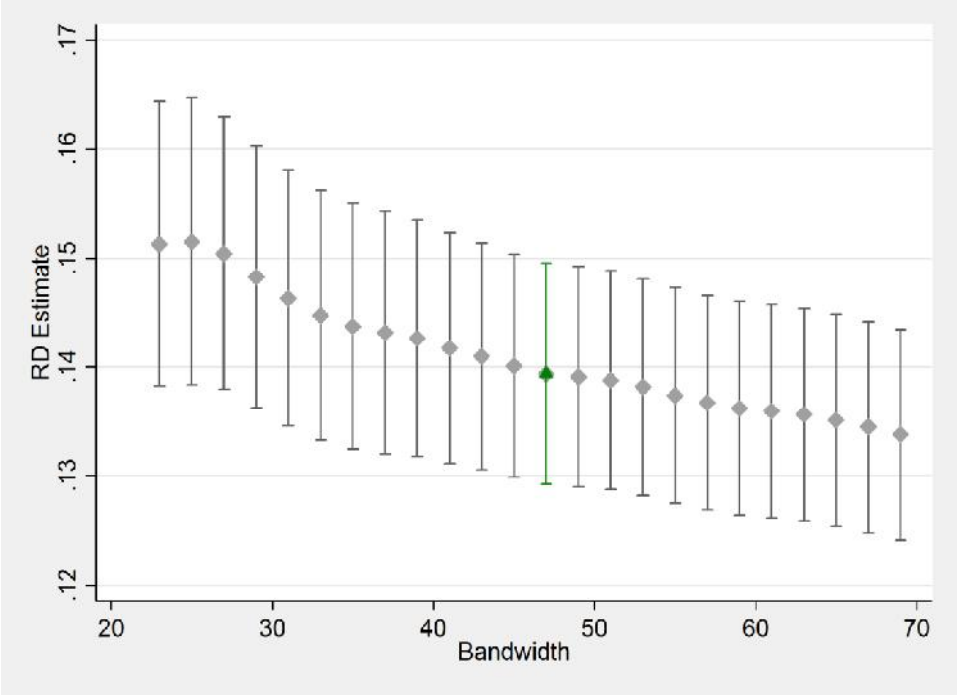
Notes: This figure reports the 2SLS estimates of the heterogeneous effect of staying in school till age 16 on mean BMI, separated by the tercile of the individual's BMI score. These 2SLS estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 19: The Effect of Staying in School till Age 16 on Diabetes by BMI Score Tercile



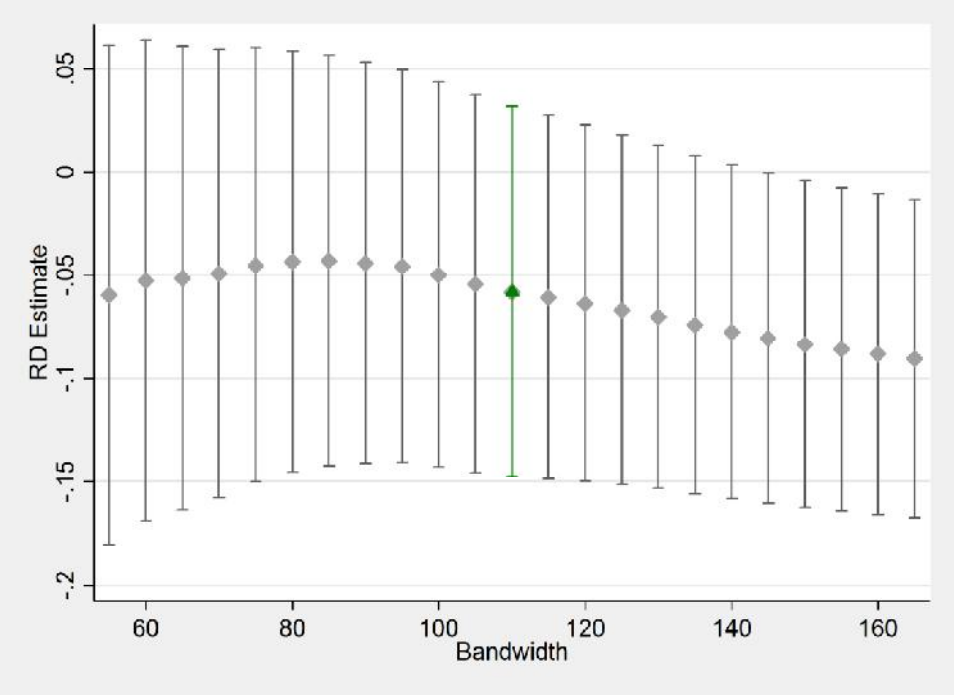
Notes: This figure reports the 2SLS estimates of the heterogeneous effect of staying in school till age 16 on diabetes risk, separated by the tercile of the individual's BMI score. These 2SLS estimates come from a local linear framework with a triangle kernel where the bandwidth of the kernel is selected by Imbens and Kalyanaraman (2012). The linear trend is allowed to vary before and after the discontinuity. Controls in the marked columns include sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 20: Bandwidth Analysis of First Stage



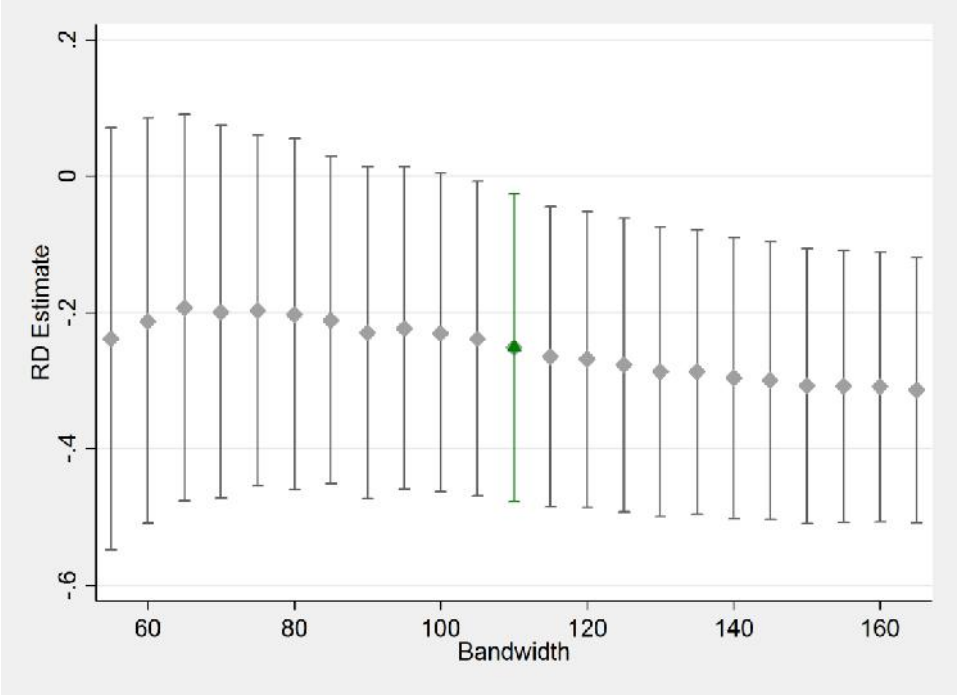
Notes: This figure displays the effect of the 1972 ROSLA on the fraction of individuals staying in school till age 16 using a variety of bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth selected by Imbens and Kalyanaraman (2012). The middle point and confidence interval shaded in green correspond to the optimal bandwidth used in the main analyses of this paper. These estimates come from a local linear framework with a triangle kernel. In all specifications, controls are included for a linear trend that is allowed to vary before and after the discontinuity, sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 21: Bandwidth Analysis of Reduced-Form (BMI)



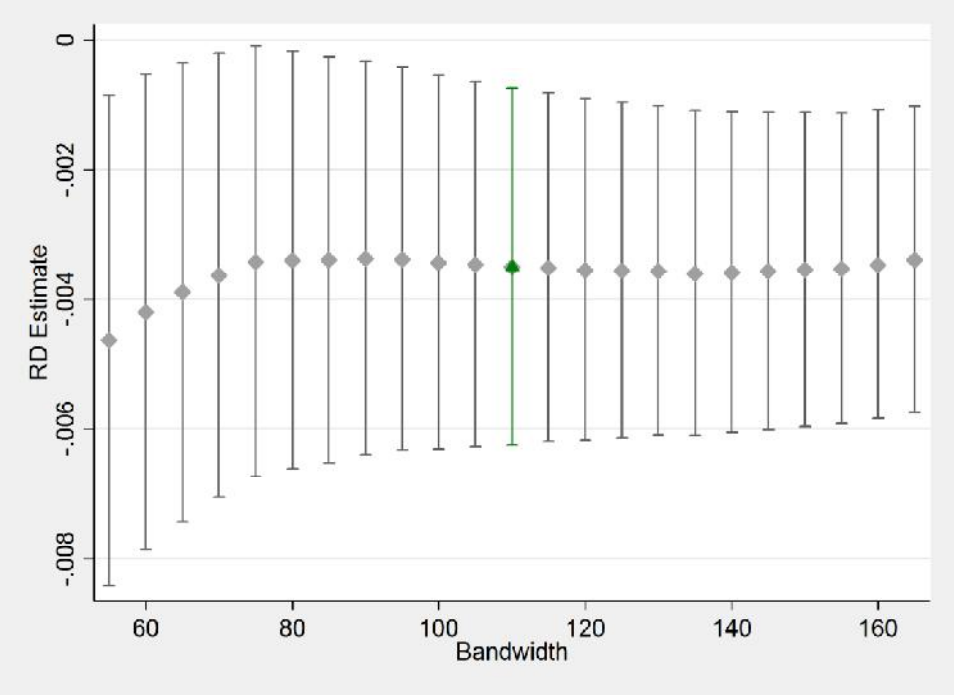
Notes: This figure displays the effect of the 1972 ROSLA on mean BMI using a variety of bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth selected by Imbens and Kalyanaraman (2012). The middle point and confidence interval shaded in green correspond to the optimal bandwidth used in the main analyses of this paper. These estimates come from a local linear framework with a triangle kernel. In all specifications, controls are included for a linear trend that is allowed to vary before and after the discontinuity, sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 22: Bandwidth Analysis of Reduced-Form (90th Percentile of BMI)



Notes: This figure displays the effect of the 1972 ROSLA on the 90th percentile of BMI using a variety of bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth selected by Imbens and Kalyanaraman (2012). The middle point and confidence interval shaded in green correspond to the optimal bandwidth used in the main analyses of this paper. These estimates come from a local linear framework with a triangle kernel. In all specifications, controls are included for a linear trend that is allowed to vary before and after the discontinuity, sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.

Figure 23: Bandwidth Analysis of Reduced-Form (Diabetes)



Notes: This figure displays the effect of the 1972 ROSLA on the fraction of individuals who reported having diabetes using a variety of bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth selected by Imbens and Kalyanaraman (2012). The middle point and confidence interval shaded in green correspond to the optimal bandwidth used in the main analyses of this paper. These estimates come from a local linear framework with a triangle kernel. In all specifications, controls are included for a linear trend that is allowed to vary before and after the discontinuity, sex, genetic scores for BMI and educational attainment, an indicator for being genotyped, fifteen principal components of the genetic matrix, month of birth fixed effects, county of birth fixed effects, and a quadratic of latitude and longitude at birth. Standard errors are clustered by month of birth.