DISTRIBUTIONAL EFFECTS OF EDUCATION ON HEALTH

Silvia H. Barcellos, Leandro S. Carvalho,

and Patrick Turley*

This paper studies distributional effects of education on health. In 1972 England, Scotland, and Wales raised their minimum school-leaving age from 15 to 16 for students born after September 1, 1957. Using a regression discontinuity design and objective health measures for a quarter million individuals, we find that education reduced body size, improved lung function, and increased blood pressure in middle age. The reduction in body size was concentrated at the upper tail of the distribution with a 7.5 percentage point reduction in obesity. The increase in blood pressure was concentrated at the lower tail of the distribution with no effect on hypertension. JEL codes: I10, I20.

Studies about the causal effect of education on health have focused on the effects of education on *average health* (e.g., Lleras-Muney 2005; Clark and Royer 2013). Average treatment effects, however, provide limited policy guidance. Small effects on average health may conceal larger effects on particular parts of the health distribution while a given average treatment effect may have different consequences for disease prevalence if effects are concentrated at the bottom or at the top of the health distribution. Different policy lessons follow depending on the exact parameter of policy interest, on the weight placed on distributional concerns, and on whether improved policy targeting is possible.[1]

---

[1] Recent studies have illustrated the importance of moving beyond mean effects when analyzing the impact of safety net programs (e.g. Black et al. 2003; Bitler, Gelbach and Hoynes 2006), social experiments (e.g. Heckman, Smith and Clements 1997; Firpo 2007; Djebbari and Smith 2008), and early childhood interventions (e.g. Bitler, Hoynes and Domina 2016).

This paper uses a regression discontinuity design to study the *distributional* effects of education on health. In 1972 England, Scotland, and Wales raised their minimum school-leaving age from 15 to 16 for students born on or after September 1, 1957 (students born before this date could drop out at age 15), generating a discontinuity in the relationship between education and date of birth at the September 1, 1957 "cutoff." Previous studies exploiting this reform (or a similar reform introduced in 1947) estimated the effect of education on average health from the discontinuity in average health at the cutoff. We estimate the distributional effects from discontinuities in the marginal cumulative distribution function (CDF) of health at the cutoff.

We use data from the UK Biobank, a study that collected multiple *objective* and *continuous* measures of health between 2006 and 2010. Using standardized protocols, healthcare technicians and nurses measured the body mass index, body fat percentage, waist and hip circumferences, lung function, and blood pressure of more than a quarter of a million individuals born in England, Scotland, and Wales between September 1, 1947 and August 31, 1967. Having this number of objective health measures for such a large sample is rare and gives us enough power to estimate distributional treatment effects. To ease concerns regarding multiple hypotheses testing, we focus our analyses on three health indices constructed from the multiple objective measures available: body size, lung function, and blood pressure.

We begin our analysis by estimating the effect of education on *average* health, finding only suggestive evidence of an impact. Specifically, staying in school until age 16 may improve one dimension of health—body size decreases on average by 0.15 of a standard deviation—while *worsening* another dimension of health—blood pressure *increases* on average by 0.15 of a standard deviation.[2] These point estimate are, however, only significant at the 10% level. We also find a marginally significant improvement in lung function that loses statistical significance once controls are added.

If the effects of education on health are heterogeneous, then a distributional test may be better powered than a test of difference in means. This would be the case, for example, if the effects of education on health are concentrated at particular parts of the health distribution (e.g., we may not expect health improvements for individuals who would have been healthy even in the absence of the additional year of schooling).

---

[2] In our sample there are virtually no participants with low blood pressure or who are underweight.

There are three main takeaways from the distributional analysis. First, it confirms that staying in school until age 16 reduces body size and improves lung function, while also increasing blood pressure. Second, it reveals that these effects vary considerably along the health distribution. The effect on body size, for example, is concentrated at the upper tail of the body size distribution. Third, the effects on body size, lung function, and on blood pressure occur in different parts of their respective distributions: while the effects on body size and lung function are concentrated at the upper tail (i.e., among the least healthy), the effect on blood pressure is concentrated at the lower tail (i.e., among the most healthy).[3]

We conduct a distributional test based on Shen and Zhang (2016) to formally investigate whether these changes are statistically significant, testing differences in the lower and upper halves of the CDFs of our three health indices. The test rejects the null of equality for the upper tails of the body size and lung function distributions, and for the lower tail of the blood pressure distribution. To give a sense of how effects are concentrated, staying in school until age 16 reduces the 90th percentile of the body size distribution by 0.38 of a standard deviation—that is equal to 2.5 times the average treatment effect of –0.15 of a standard deviation.

These results illustrate the policy relevance of studying distributional effects. Even though the average treatment effects show a reduction in body size and an increase in blood pressure of identical magnitudes (0.15 of a standard deviation), a policy maker may wish to trade-off these effects based on which parts of the respective distributions are affected. Because the reductions in body mass index (BMI) occur at the right tail of the BMI distribution, staying in school until age 16 reduces obesity rates (i.e., BMI above 30) by 7.5 percentage points. In contrast, the increase in blood pressure is concentrated below the clinical threshold for hypertension (i.e., diastolic blood pressure above 90 mmHg or systolic blood pressure above 140) with no consequences for hypertension prevalence.[4]

We also conduct an exploratory analysis to investigate possible channels through which education may affect health. We find that staying in school until age 16 increases socioeconomic status (SES): the more educated have higher household income, own more cars, live in higher-SES

---

[3] We do not have the power to test whether some people experienced both a reduction in body size and an increase in blood pressure. However, given that there is a moderate positive correlation between these two outcomes ($\rho = 0.3$), it is unlikely (though possible) that those who experienced reductions in body size were the same individuals experiencing increases in blood pressure.

[4] There could still be harmful health consequences, as observational studies have linked increases in blood pressure in the range documented here to increased ischemic heart disease and stroke mortality risk (Lewington et al. 2002).

neighborhoods, and work in higher-SES occupations. The additional schooling also improves the diet of the more educated, who eat less fat and less saturated fat. We find, however, no effects of education on smoking and on physical activity. While improvements in SES and diet may explain the reduction in body size, we are not able to uncover in our data channels through which education may increase blood pressure. We find no effects of education on hypertension diagnosis or utilization of blood pressure medication, for example. We speculate that, by changing the types of occupations and careers individuals have, education might have an effect on job responsibilities, expectations, and work-related stress. This would be consistent with evidence that academically successful African Americans in the U.S. have higher biomarkers related to cardiometabolic risk (i.e. blood pressure and stress hormones) than other groups (Brody et al. 2013, Miller et al. 2015, Chen et al. 2015). However, we are not able to offer direct evidence of this channel.

There are a number of studies that have exploited changes in compulsory schooling laws to study the causal effect of education on average health (e.g., Lleras-Muney 2005; Albouy and Lequien 2009; Silles 2009; Powdthavee 2010; Kemptner et al. 2011; Clark & Royer 2013; Jürges et al. 2013; Davies et al. 2017).[5] This paper adds to this literature in the following ways. First, to the best of our knowledge, this is the first paper to estimate the distributional effects of education on health using a quasi-experimental approach.[6] Second, we provide evidence about the importance of conceptualizing health as a multidimensional construct. We show that education improves body size and lung function while also increasing blood pressure. Finally, we consider only high-quality, objective and continuous measures of health. Previous research shows that discrepancies between subjective (e.g. self-reported hypertension) and objective measures of health (e.g. objectively measured hypertension) vary with socioeconomic status (e.g. Johnston et al. 2009).

Our results can be more easily compared to Clark and Royer (2010) [CR, hereafter] and concurrent work by Davies et al. (2017) [DDSBW, hereafter], both of which also use the 1972 school leaving-age reform to study the effects of education on health, albeit with a focus on the effects on average health. Our estimates of the effects on average BMI and average diastolic and

---

[5] This literature has produced mixed evidence, resulting in no consensus on whether there is a causal effect of education on health. Grossman (2015) reviews this literature and concludes that "there is enough conflicting evidence to warrant more research on the question of whether more schooling does in fact cause better health outcomes."

[6] Conti, Heckman and Urzua (2011) have used a structural approach to estimate the distributional effects of education on health.

systolic blood pressures lie within the 95% confidence intervals of CR's and DDSBW's estimates (CR do not look at the effects on systolic blood pressure). The only exception being DDSBW's estimate of the effect on BMI: Although our estimate and DDWBW's estimate have the same sign, our estimate is smaller than the lower bound of DDSBW's estimate – see Appendix Table D3. Since DDSBW also use UK Biobank data this difference is probably due to the slightly different specification (bandwidth and polynomial choice) and sample selection (they do not include Wales and Scotland). Neither of these two papers look at the effects on lung function. Our main contribution relative to CR and DDSBW is to go beyond mean effects and estimate the distributional effects of education on health, focusing on objective and continuous health measures.[7]

The paper is structured as follows. Section 1 discusses the 1972 raising of the school leaving-age reform and the data. In Section 2 we present the effects of the reform on education and the effects on average health. Section 3 discusses the methods used to estimate the distributional effects with results shown in Section 4. Section 5 presents suggestive evidence on mechanisms and Section 6 concludes.

1. **Background and Data**

A. The 1972 Raising of the School Leaving Age

The British compulsory schooling laws specify the maximum age by which children must start school and the minimum age at which they can leave school. In this paper, we exploit the 1972 Raising of School Leaving Age (ROSLA) legislation, which increased the minimum school-leaving age from 15 to 16 years of age in England, Scotland, and Wales. These laws and their implementation have been extensively documented in other studies (see Clark and Royer 2010, 2013) so we only include a brief summary of its main features here.

The UK's 1944 Education Act raised the minimum school-leaving age from 14 to 15 years of age in England, Wales, and Scotland and gave the Minister of Education the power to further raise

---

[7] Both studies also test for the effect of the reform on overweight, obesity, and/or hypertension, which is equivalent to a distributional test at a single point. Similar to us, CR do not find an effects on probability of being overweight or hypertensive. They also do not find a significant reduction in obesity, but that might be due to lack of power (at 18K observations their sample is 7.5 times smaller than ours is). Similar to us, DDSBW do not find that the reform affected hypertension (they use a self-reported measure). DDSBW do not analyze obesity/overweight.

it to 16 years when conditions allowed. The Minister did so in January 1972 for Scotland (Statutory Instrument No. 59)[8] and in March 1972 for England and Wales (Statutory Instrument No. 444)[9]. Both changes took effect in September 1, 1972, implying that those who were 15 or younger before that date (born on September 1, 1957 or later) had to stay in school until at least age 16 in the three countries (hereafter, we use the term "stayed in school until age 16" to refer to those who stayed in school until *at least* age 16). Infrastructure investments, such as school building to absorb the additional students, preceded the 1972 ROSLA but key elements of the school system did not change with the policy.

Figure 1, which displays the fraction of study participants who stayed in school until age 16 (y-axis) by quarter of birth (x-axis), shows that the policy generated a discontinuous relationship between these two variables. There is a large jump at the September 1, 1957 cutoff marked by the vertical dashed line.[10] We estimate that the policy increased the fraction of student participants who stayed in school until age 16 by 15 percentage points – see Table 1.[11]

---

[8] http://www.legislation.gov.uk/uksi/1972/59/pdfs/uksi_19720059_en.pdf
[9] http://www.legislation.gov.uk/uksi/1972/444/pdfs/uksi_19720444_en.pdf
[10] Note the cyclical drop in the fraction of students staying in school until age 16 or later, corresponding to the fourth quarter of each year. Those born during the summer months could in practice drop out at age 15 even after the 1972 ROSLA, since the law required students to be 16 by the start of the next school year.
[11] Other studies using nationally representative data estimate that the policy increased the fraction of students staying in school until age 16 by approximately 25 percentage points (e.g., Clark and Royer 2013). This difference may be due to the composition of the UK Biobank sample, which is more educated than the overall population.

Figure 1: Fraction Staying in School until Age 16 by Quarter of Birth
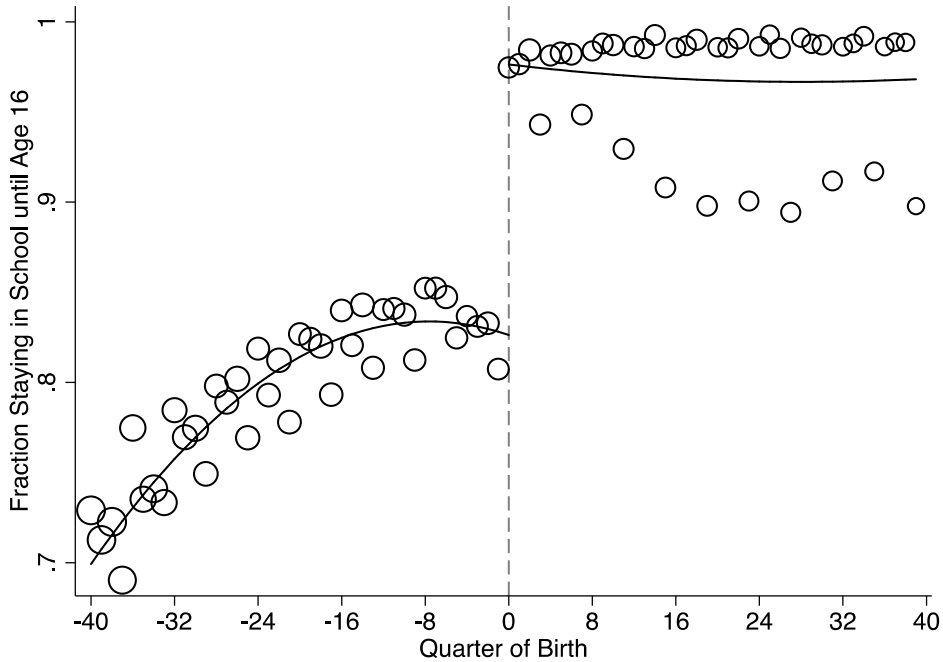
Notes: The figure shows the fraction of study participants who stayed in school until age 16 by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

B. Data

We 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 9.2 million people between the ages of 40 and 69 who were registered with the NHS and lived up to about 25 miles from one of 22 study assessment centers distributed throughout the UK (Allen et al. 2012) – see Appendix Figure D1.[12] The sample is formed by 503,325 individuals who agreed to participate (i.e. a response rate of 5.47%). Although the sample is not nationally representative, our estimates have internal validity because there is no differential selection on the two sides of the September 1, 1957 cutoff – see Appendix B.

---

[12] The NHS has contact details for an estimated 98% of the UK population.

Study participants went through an assessment that comprised a self-completed touch-screen questionnaire; a brief computer-assisted interview; physical and functional measures; and collection of blood, urine, and saliva. The collection of physical measures, which included among others anthropometrics, spirometry, and blood pressure, was standardized across centers and was conducted by trained nurses or healthcare practitioners. About 100,000 participants also wore accelerometers that recorded physical activity for 7 days. Every participant was genotyped.

In this paper, we focus on objective and continuous measures of health. Continuous measures because we are interested in studying how education affects the distribution of health. Objective measures because research shows that discrepancies between subjective (e.g. self-reported hypertension) and objective measures of health (e.g. objectively measured hypertension) vary with socioeconomic status (e.g. Johnston et al. 2009). We restrict ourselves to three dimensions of health (that satisfy the two criteria above and) that can be arguably affected by education: *body size*, *lung function*, and *blood pressure*. The UK Biobank has multiple measures of each of these health dimensions.

Next, we describe how each of these health dimensions is measured in the data.

Body Size

We use three measures of body size: body mass index (BMI), body fat percentage, and waist-hip ratio.[13] A bioimpedance analyzer was used to calculate body fat percentage. This device passes a low electrical current through the body. Water conducts electricity. While fat contains very little water, muscle contains 70% water. The bioimpedance analyzer calculates body fat from the speed of the current: The slower the signal travels, the greater the fat content. Weight and body size can be affected by education through diet and physical exercise.

Lung Function

A spirometry test was conducted to measure participants' lung function. The spirometer is a small machine attached to a mouthpiece by a cable that measures the volume and speed of air after a forced exhale. Participants were asked to fill their lungs as much as possible and to blow air out as hard and as fast as possible in the mouthpiece.[14] Three parameters were measured: 1) *forced*

---

[13] The UK Biobank provides two measures of BMI: one calculated as weight in kilograms divided by height squared (in meters) and one using height and electrical impedance to quantify mass. We take the average of these two measures. The waist-hip ratio is equal to the waist circumference divided by the hip circumference.

[14] They were instructed to continue blowing until no more air came out of their lungs. Up to three attempts were allowed. The participant was allowed a third attempt if the first two blows did not satisfy the reproducibility criteria of the spirometry protocol.

expiratory volume in the first second is the amount of air exhaled during the first second; 2) *forced vital capacity* is the total amount of air exhaled during the forced breath; and 3) *peak expiratory flow* is the fastest rate of exhalation. These parameters are used to assess pulmonary conditions, such as chronic obstructive pulmonary disease and asthma. We follow DeMateis et al. (2016)'s criteria to identify acceptable expiratory maneuvers in the UK Biobank data. Valid spirometry measures are available for 79% of our sample.[15] Lung function can be affected by education through smoking and pollution.

Blood Pressure

Two measurements were taken of the diastolic and systolic blood pressures of each study participant. We use the average of these two measurements. Blood pressure may be affected by education through diet, physical exercise, medication, and stress.

Summary Indices

In order to reduce the number of outcomes and partly address concerns about multiple hypothesis testing, we construct for each health dimension a summary index that is a weighted average of the different outcomes measuring that dimension:

1. *Body size*: body mass index, waist-to-hip ratio, and body fat percentage;
2. *Lung function*: forced expiratory volume in the first second, forced vital capacity, and peak expiratory flow;
3. *Blood pressure*: diastolic and systolic blood pressures.

First, each measure is standardized separately by gender, using as a reference those born in the 12 months before September 1, 1957. We then follow the procedure proposed by Anderson (2008), weighing the measures by their variance-covariance matrix. The weights are calculated to maximize the amount of information captured in the index. Finally, we construct a fourth "summary index" that is a summary of the body size, the lung function, and the blood pressure indices, using the same weighting procedure. We construct all four indices so that a higher number corresponds to worse health.

The correlation between the body size and lung function indices is 0.20. The correlation between the body size and the blood pressure indices is 0.30. The correlation between the lung

---

[15] Appendix Figure C5 and Appendix Table C1 show that participants born before and after September 1957 are equally likely to have valid spirometry measures.

function and the blood pressure indices is 0.10. The correlations between the summary index and the body size, lung function, and blood pressure indices are respectively 0.69, 0.67, and 0.68.

2. Mean Effects

A. Effects of the Compulsory Schooling Change on Education

We use a regression discontinuity design (RDD) to estimate the "first stage", i.e., the effect of the 1972 ROSLA on education. In particular, we estimate the following regression:

$$Educ_i = a_0 + a_1 Post_i + f(DoB_i) + x_i' a_2 + \varepsilon_i, \tag{1}$$

where $Educ_i$ is a measure of the educational attainment of individual i; $Post_i$ is 1 if individual i was born on or after September 1, 1957 (and 0 otherwise); DoB_i is individual i's date of birth; and the vector x_i contains predetermined characteristics. Date of birth is measured in days relative to the cutoff, such that $DoB = 0$ for someone born on September 1, 1957. The function $f(\cdot)$ captures birth cohort trends in educational attainment, which are allowed to differ on either side of the September 1, 1957 cutoff. The coefficient a_1 gives the effect of the 1972 ROSLA on educational attainment.[16]

We adopt a global polynomial approach in our estimation (see Lee and Lemieux 2010). We restrict the data to study participants born within 10 years of September 1957 – that is, born between September 1, 1947 and August 31, 1967 – and use a quadratic polynomial in date of birth to capture cohort trends (i.e., function $f(\cdot)$ in equation (1)). In Appendix B we show our main results are robust to the choice of bandwidth and to the use of linear trends.[17] We use triangular kernel weights that give greater weight to study participants born closer to the cutoff. The set of predetermined characteristics include gender, age in days (at the time of the baseline assessment) and age squared, dummies for ethnicity, dummies for country of birth, and dummies for calendar month of birth (to control for seasonality).[18]

---

[16] The inclusion of predetermined controls in equation (1) is not needed for identification but can improve the precision of estimates.

[17] Gelman and Imbens (2016) caution against the use of higher order polynomials (higher than 2) in RDD.

[18] Because participants were surveyed for the baseline assessment between 2006 and 2010, date of birth and age are not perfectly collinear.

Table 1 shows estimates of effects of the 1972 ROSLA on education. Each cell reports results from a separate ordinary least squares estimation of (1), where we vary the dependent variable (listed in the column) and whether the predetermined characteristics are included as controls. The table shows the coefficient on the indicator variable for being born on or after September 1, 1957, a_1, and the mean of the dependent variable among those born in the 12 months before September 1, 1957. Robust standard errors are reported between brackets.

We estimate that the 1972 ROSLA increased the fraction of study participants staying in school until age 16 by 14-15 percentage points, an estimate significant at the 1% significance level.[19]

Table 1: Effects on Education

	Left school at age ≥ 16		No qualification		CSE	
Post	0.150	0.139	-0.048	-0.050	0.059	0.070
	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.005]***	[0.005]***
Controls?	No	Yes	No	Yes	No	Yes
Mean of Y	0.827		0.113		0.205	

	O-level		A-level		College degree	
Post	0.038	0.035	0.016	0.015	-0.003	-0.005
	[0.006]***	[0.006]***	[0.006]***	[0.006]**	[0.006]	[0.006]
Controls?	No	Yes	No	Yes	No	Yes
Mean of Y	0.513		0.325		0.368	

Notes: The table shows the effects of the school reform on education. Each cell corresponds to a separate regression. We report the coefficient on the indicator variable for being born on or after September 1, 1957 (i.e., "Post"). The dependent variable mean in the bottom row is the weighted mean among those born in the 12 months before September 1, 1957. Controls include male, age in days and age squared, dummies for calendar month of birth, dummies for ethnicity, and dummies for country of birth. Robust standard errors. $N = 271,082$ for "*Stayed in school until 16*" and $N = 268,551$ for all other outcomes.

---

[19] Estimates of the effect of the 1972 school-leaving age reform on staying in school until age 17 or later are an order of magnitude smaller than the effect on staying in school till age 16 and are generally not robust to the inclusion of controls – see Appendix Figure D2 and Appendix Table D1.

One may worry that these students forced to stay in school an extra year did not learn much if they did not put effort into it. The evidence does not support this hypothesis. By the 70's high schools offered a series of two-year courses that ran through grades nine and ten and required students to take exams at the end of grade ten (the grade they are typically in by age 16): Certificate of Secondary Education (CSE) or a General Certificate of Education (GCE) Ordinary Level (also known as an O-level). By compelling students to stay in school until grade ten, the 1972 ROSLA gave students an incentive to complete these courses and get these qualifications, which are valued in the labor market (Dickson and Smith 2011).

Figure 2: Fraction with a CSE or O-level Qualification by Quarter of Birth
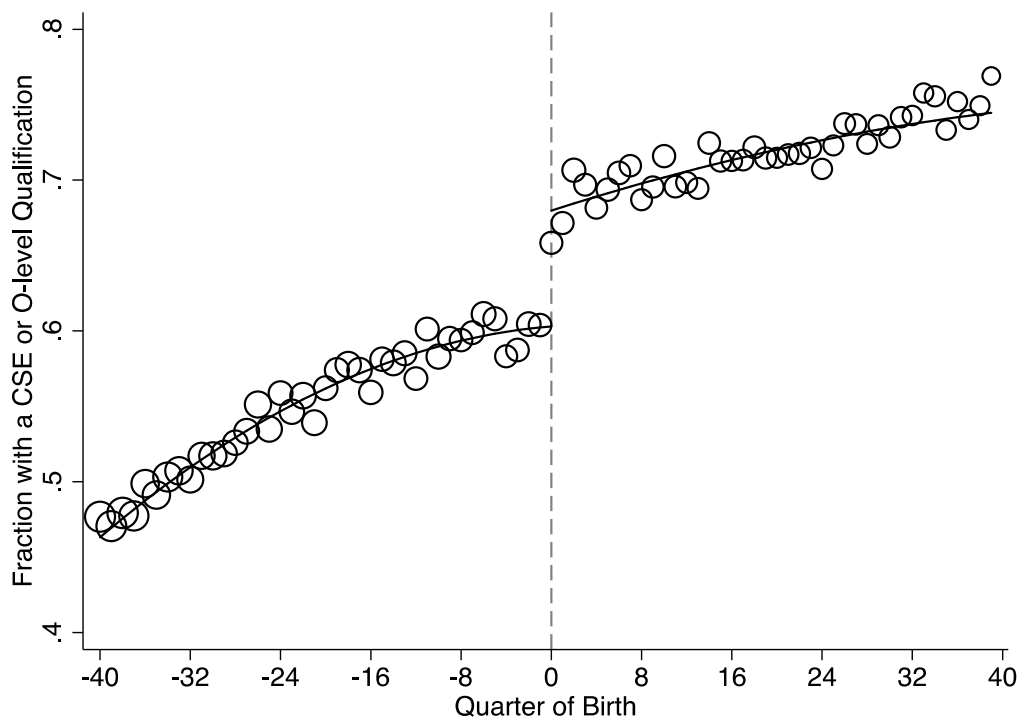
Notes: The figure shows the fraction of study participants with a CSE or O-level qualification by quarter of birth. The dashed vertical line marks the introduction of the policy. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 268,551$.

Figure 2 shows that the policy generated a discontinuous increase in the fraction of study participants with these qualifications. In Table 1 we estimate that the policy increased the fraction of study participants with a CSE by 6-7 percentage points and the fraction with an O-level by 3-4

percentage points. Interestingly, the fraction with an A-level, an exam typically taken at age 18 and used for college admissions, increased by 1-2 percentage points. The fraction without any formal qualification dropped by 5 percentage points. All of these estimates are statistically significant at 1%. We find no effect of the policy on having a college degree.

B.   Effects on Average Health

We now turn to the effects of the 1972 ROSLA on average health. We are interested in the relationship between health and education:

$$Health_i = \beta_0 + \beta_1 Educ16_i + g(DoB_i) + x_i'\boldsymbol{\beta}_2 + u_i, \tag{2}$$

where $Health_i$ is a health measure for individual i. $Educ16_i$, an indicator for whether individual i stayed in school until age 16, is our endogenous measure of education. The function $g(\cdot)$ captures birth cohort trends in health and is allowed to differ on either side of the September 1, 1957 cutoff. We substitute (1) into (2) to get the "reduced-form" effect of the 1972 ROSLA on *average* health:

$$Health_i = \gamma_0 + \gamma_1 Post_i + j(DoB_i) + x_i'\boldsymbol{\gamma}_2 + v_i. \tag{3}$$

The coefficient γ_1 gives the effect of the school leaving-age reform on average health. The RDD identifying assumption is that, in the absence of the reform, our outcomes of interest would have been smooth across the September 1, 1957 threshold. This assumption is violated if determinants of health are discontinuous at the cutoff (Lee 2008). In Appendix A we partially test for such violations by investigating whether the average (or the cumulative distribution function) of predetermined characteristics, such as gender and place of birth, are discontinuous around September 1, 1957. We find no evidence that this is the case, which strengthens our confidence that the RDD results provide unbiased estimates of the causal effects of education on the health of UK Biobank participants.

We estimate the causal effect of staying in school until age 16 on average health, β_1, through two stages least squares (2SLS), using the indicator for being born on or after September 1, 1957 (i.e., $Post_i$) to instrument for staying in school until age 16 (i.e., $Educ16_i$) in equation (2). We adopt the same specifications used to estimate the effects on education (see section 2.A), namely:

a global polynomial approach, 10-year bandwidths, quadratic polynomials to capture birth cohort trends, triangular kernel weights, and the same set of controls. Appendix B shows our results are robust to linear cohort trends and smaller bandwidths.

Figure 3 examines the effects of the 1972 compulsory schooling change on average health. The graphs show average health (y-axis) by quarter of birth (x-axis), where health is measured by the four health indices: the body size index (top left); the lung function index (top right); the blood pressure index (bottom left); and the summary index (bottom right).

Table 2 displays corresponding regression estimates. The first rows show the coefficients on the indicator variable for being born on or after September 1, 1957, γ_1 in equation (3), from reduced-form estimates. The third row shows the coefficients on staying in school until age 16 from 2SLS estimates, β_1 in equation (2), where the indicator variable for being born on or after September 1, 1957 is used to instrument for staying in school until age 16. Again, the health indices were constructed such that higher values correspond to worse health.

Overall Figure 3 suggests education may lead to small average improvements in health, with minor discontinuous decreases in the body size, lung function, and summary indices at the cutoff. One noteworthy exception is blood pressure. There is a discontinuous *increase* in the blood pressure index at the cutoff, suggesting that education may *worsen* this particular dimension of health.

Table 2 shows that the effects on body size and blood pressure are statistically significant at the 10% significance level. The 2SLS point estimates imply that staying in school until age 16 decreases the body size, the lung function, and the summary indices respectively by 0.15-0.16, 0.17, and 0.12 of a standard deviation. At the same time, staying in school until age 16 *increases* the blood pressure index by 0.15 of a standard deviation.

Table 2: Effects on Average Health

	Body Size		Lung Function		Blood Pressure		Summary	
Reduced-form								
Post	-0.023	-0.023	-0.024	-0.022	0.023	0.021	-0.016	-0.016
	[0.013]*	[0.013]*	[0.014]*	[0.014]	[0.013]*	[0.013]*	[0.014]	[0.014]
Two stages least squares								
Stayed in school until 16	-0.154	-0.163	-0.175	-0.174	0.151	0.151	-0.120	-0.125
	[0.083]*	[0.091]*	[0.103]*	[0.112]	[0.084]*	[0.091]*	[0.103]	[0.112]
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
N Observations	266,525	266,525	215,536	215,536	270,647	270,647	212,689	212,689

Notes: The table shows the effects on average health. The first two rows show reduced-form effects of the 1972 Raising of the School Leaving Age. The last two rows show two stages least squares estimates of the effect of staying in school until age 16 obtained by using an indicator for being born on or after September 1, 1957 to instrument for staying in school until age 16. Robust standard errors. Controls include male, age in days and age squared, dummies for calendar month of birth, dummies for ethnicity, and dummies for country of birth.

Figure 3: Effects on Average Health
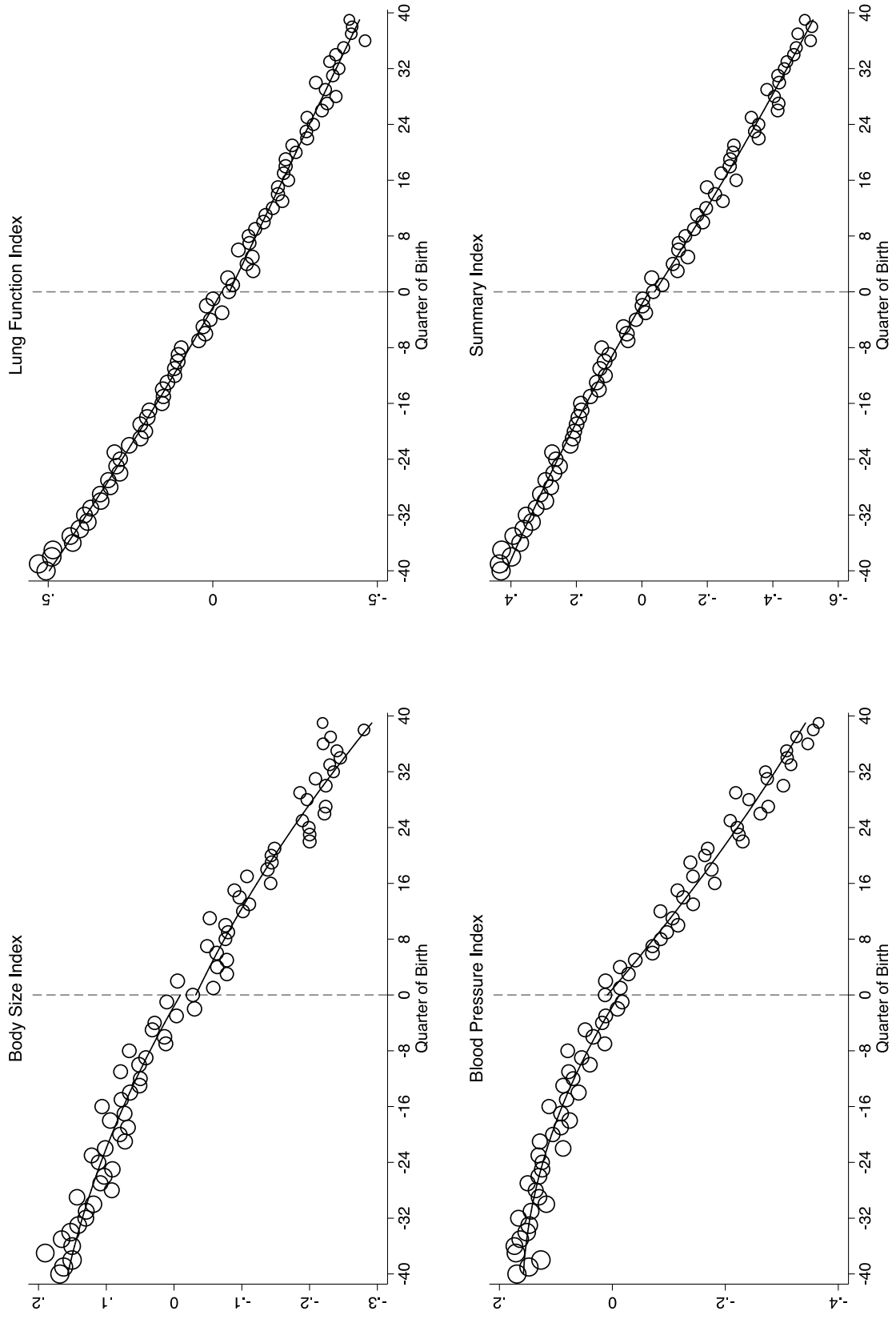
Notes: These figures show average health by quarter of birth. See Table 2 for number of observations.

These results point to the importance of treating health as multi-dimensional and considering the effects of education on different dimensions separately. Focusing on the analysis of summary measures of health can lead to misleading conclusions of no health impact if effects going on opposite directions cancel out, as is the case in Table 2.

Even though the effects on average health are informative, it may conceal larger effects on particular parts of the health distribution with important policy implications. In the next section, we describe the methods we use to estimate how education affects the distribution of health.

3. Methods for Distributional Effects Estimates

In Section 2.B, we estimated the effect of education on *average health* (of compliers) by investigating if there was a discontinuity in the relationship between *average health* and date of birth at the September 1, 1957 cutoff. Here we estimate the effect of education on the *health distribution* (of compliers) by investigating if, at the September 1, 1957 cutoff, there is a discontinuity in the relationship between the *cumulative distribution function (CDF) of health* and date of birth.[20] We are interested in estimating the local distributional treatment effect (LDTE) for the compliers. Let the *pre-reform CDF* be the CDF for the compliers in the limit when date of birth is converging to September 1, 1957 from the left (i.e., $DoB < 0$):

$$F_{pre}(h) = \lim_{DoB \to 0^-} \Pr(Health \le h | DoB).$$

Similarly, the *post-reform CDF* is defined as the CDF for the compliers in the limit when date of birth is converging to September 1, 1957 from the right (i.e., $DoB > 0$):

$$F_{post}(h) = \lim_{DoB \to 0^+} \Pr(Health \le h | DoB).$$

The LDTE, which is the discontinuity in the CDF at September 1, 1957, is estimated as the difference between $F_{post}(h)$ and $F_{pre}(h)$ for any given h:

$$\mu(h) = F_{post}(h) - F_{pre}(h). \tag{4}$$

In practice, we estimate $F_{pre}(\cdot)$ using the following equation:

---

[20] The RDD identifies differences in the marginal distributions of cohorts affected and unaffected by the reform. Stronger assumptions (such as rank preservation) would be needed to estimate the distribution of treatment effects.

$$\Pr(Health_i \leq h) = \delta_0(h) + k(DoB_i|h) + \xi_i(h) \tag{5}$$

while restricting the sample to respondents born before September 1, 1957 and who left school at age 15 or younger (i.e., "the compliers"). $\Pr(Health_i \leq h)$ is the probability that the health index for individual i is smaller or equal to h. The function $k(\cdot|h)$ captures *pre-reform* birth cohort trends. Since there are very few never-takers in our sample (i.e., individuals who would leave school before age 16 whether they were born before or after September 1, 1957), by restricting the sample this way we can closely represent the compliers born before September 1, 1957. Therefore, the function $\delta_0(h)$ provides an estimate of $F_{pre}(h)$.

Next, we estimate the LDTE – $\mu(h)$ in equation (4) – by running the following set of regressions:[21]

$$\Pr(Health_i \leq h) = \theta_0(h) + \theta_1(h)Educ_i + l(DoB_i|h) + \boldsymbol{x}_i'\boldsymbol{\theta}_2(h) + \varepsilon_i(h), \tag{6}$$

where the birth cohort trends $l(\cdot|h)$ are allowed to differ on either side of the cutoff date. We estimate (6) through 2SLS using the indicator variable for being born on or after September 1, 1957 ($Post_i$) to instrument for staying in school until age 16 ($Educ_i$). The function $\theta_1(h)$ provides estimates of $\mu(h)$.

Inference based on the standard errors generated by 2SLS estimates of (6) is problematic because it leads to a large number of highly correlated statistical tests, raising concerns about multiple hypothesis testing. We, therefore, use a single distributional test based on Shen and Zhang (2016) to formally investigate whether education changes the distribution of health. Our test compares the pre- and post-reform CDFs *of the whole population*. Under the assumptions of Shen and Zhang (2016), however, any discontinuity in the CDF of the population necessarily implies that there is a discontinuity in the CDF of compliers. This test is therefore based on the reduced-form specification:

$$\Pr(Health_i \leq h) = \kappa_0(h) + \kappa_1(h)Post_i + l(DoB_i|h) + \boldsymbol{x}_i'\boldsymbol{\kappa}_2(h) + \eta_i(h). \tag{7}$$

---

[21] The CDF approach used here is closely related to a quantile IV approach. The CDF approach gives us the vertical difference of CDFs while the quantile approach gives us the horizontal difference, therefore either approach lead us to the same substantive conclusions. We opted to present the CDF approach because it is the framework used by Shen and Zhang 2016, whose results we use in our distributional tests.

The basis of our test is that—under the null hypothesis of no effect on the health distribution—the function of estimates $\hat{\kappa}_1[h(\tau)]$, where $h(\tau)$ is the value corresponding to the τth quantile of $Health_i$, is a Brownian bridge (Shen and Zhang 2016).[22] We can therefore perform an Anderson-Darling test (Anderson and Darling 1952) using the following weighted integral as our test statistic:[23]

$$T = \int_0^1 \frac{\hat{\kappa}_1[h(\tau)]^2}{\tau(1-\tau)} d\tau. \tag{8}$$

Average treatment effects may not detect a causal effect of education on health if education has a larger effect on the tails of the health distribution. We chose the Anderson-Darling test because it is uniformly powered for the whole range $\tau \in [0, 1]$ (Stephens 1974). To test for differences in the lower half of the health distribution, we use a modified version of (8), integrating only from zero to 0.5. Similarly, we test differences in the upper half by integrating from 0.5 to 1.[24]

The p-values for the test are calculated by simulation. Specifically, we generate an independent, standard normally distributed outcome for each individual (such that there is no discontinuous change in distribution at the discontinuity), and evaluate T (or the upper and lower distribution analogue) for this simulated variable. By Shen and Zhang (2016), this is equivalent to drawing from the test statistic distribution under the null. This is repeated 5,000 times. As the p-value, we report the fraction of times our simulated values of T are greater than our estimated value of T.

---

[22] More precisely, the difference in the empirical CDFs estimated in this way is a standard Brownian bridge times a scalar. See Shen and Zhang (2016) for details on calculating the scalar which allows us to transform the difference into a standard Brownian bridge.

[23] Shen and Zhang (2016) use $\max_\tau |\hat{\kappa}_1[h(\tau)]|$ as their test statistic, which corresponds to a Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test has been shown to be well-powered for deviations in the distribution near the median of the distribution, but is poorly powered to detect differences in the distribution in the tails (Stephens 1974).

[24] In practice, we calculate the integral T numerically, using the approximation

$$T \approx \sum_j \frac{1}{100} \frac{\hat{\kappa}_1[h(\tau_j)]^2}{\tau_j(1 - \tau_j)}$$

where $\{\tau_j\}$ is a set of discrete points in 0.01 unit increments. When testing the full distribution we sum from 0.01 to 0.99, inclusive. For the lower or upper portion of the distribution, we sum from 0.01 to 0.50 or 0.50 to 0.99 inclusive, respectively.

4. Distributional Effects of Education on Health

Figure 4 shows the distributional treatment effects of education on body size. It shows the pre- and post-reform CDFs of the body size index for the compliers. The former corresponds to estimates of $\delta_0(h)$ from equation (5).[25] To estimate the post-reform CDF, for any given h we add the estimate of $\theta_1(h)$ from equation (6) to the estimate of $\delta_0(h)$ from equation (5).

Figure 4 shows that education reduces body size: The post-reform CDF is shifted to the left relative to the pre-reform CDF. Importantly, the shift is not parallel; the gains are concentrated at the top of the distribution, among the least healthy. Staying in school until age 16 increases the fraction of study participants with a body size index smaller than 1 standard deviation from 77.5% to 84.4%. Similarly, the 90th percentile of the body size distribution decreases from 1.58 to 1.2 standard deviations. This effect is 2.5 times the average treatment effect (on the treated) of -0.15 standard deviations estimated in Table 2.
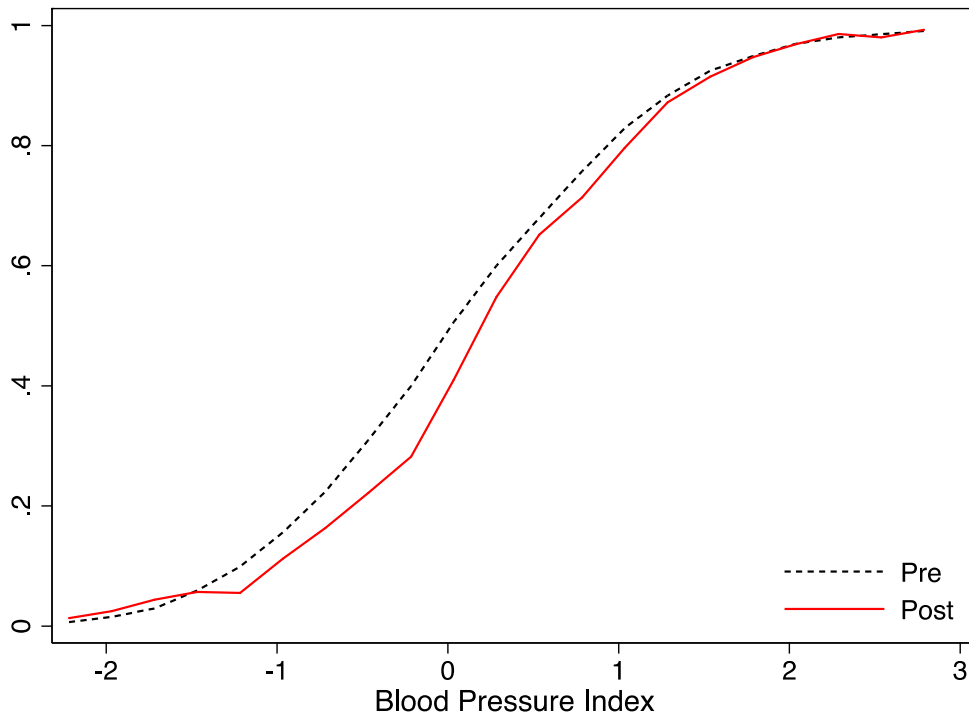
Figure 4: Distributional Effects on Body Size Index
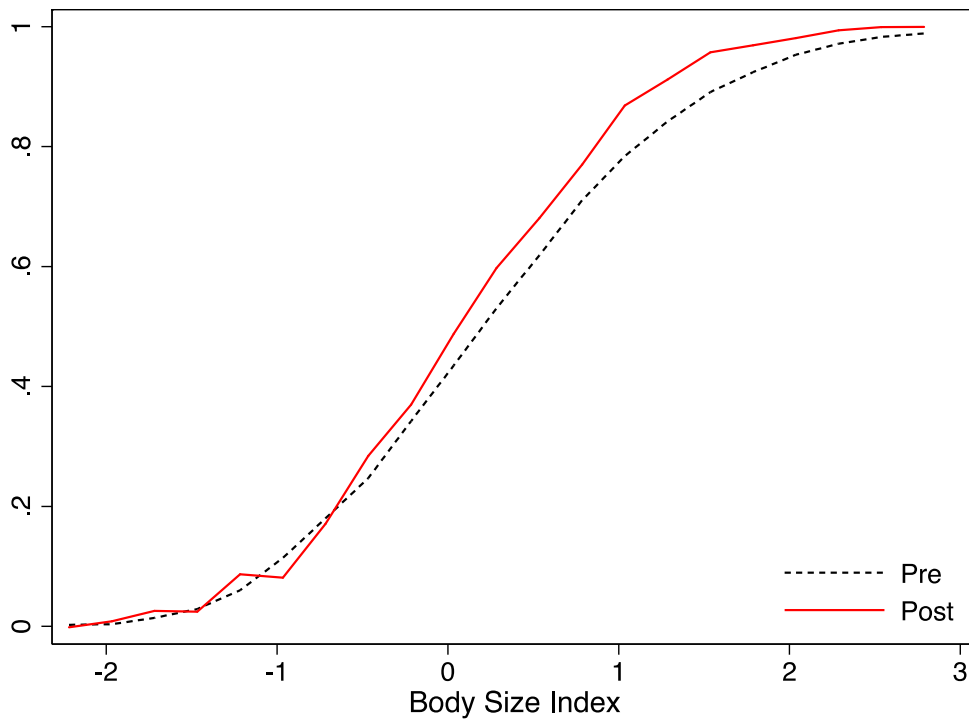
Notes: The figure shows the pre- and post-reform CDFs of the body size index for compliers. $N = 266,525$.

---

[25] Compliers are less healthy than always takers but not dramatically so. Appendix Figures D4, D5, and D6 compare the pre-reform CDFs for compliers and the whole population (both estimated using equation (5)) for our 3 indices.

Figure 5 shows that education also improves lung function: The post-reform CDF is shifted to the left relative to the pre-reform CDF. Staying in school until age 16 increases the fraction of study participants with a lung function index smaller than 1 standard deviation from 78.1% to 87.6%. Similarly, the 90th percentile of the lung function distribution decreases from 1.48 to 1.14 standard deviations. This effect is 2 times the average treatment effect (on the treated) of -0.17 standard deviations estimated in Table 2.

Figure 5: Distributional Effects on Lung Function Index
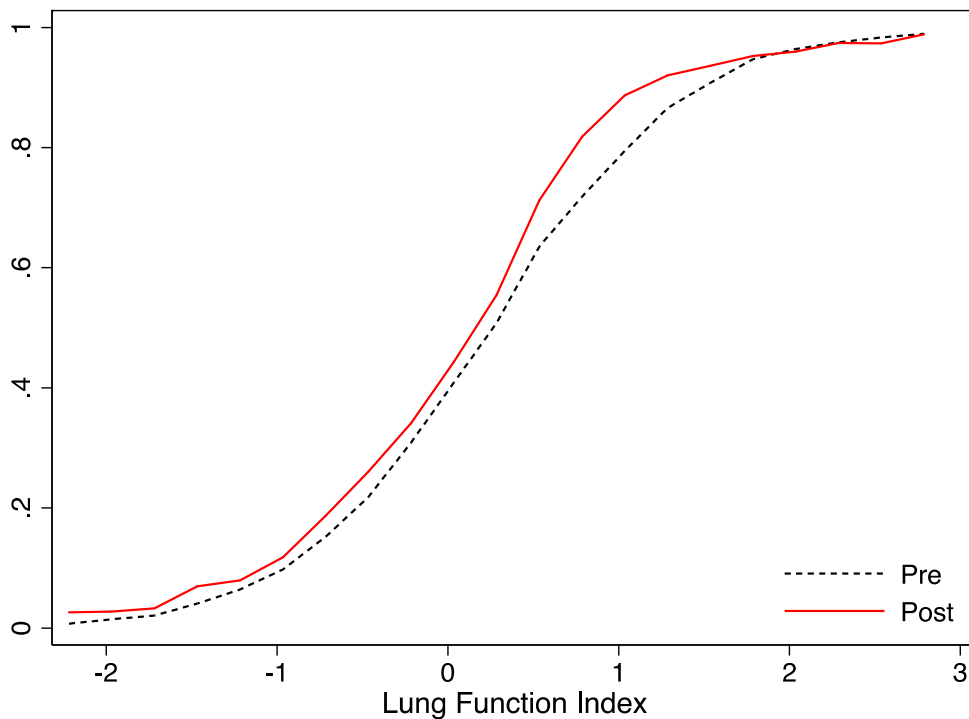
Notes: The figure shows the pre- and post-reform CDFs of the lung function index for compliers. *N* = 215,366.

While Figures 4 and 5 show that education improves body size and lung function, Figure 6 shows that education *worsens* one dimension of health: it increases blood pressure.[26] The post-reform CDF lies to the *right* of the pre-reform CDF. Staying in school until age 16 decreases the fraction of study participants with a blood pressure index smaller than 0 from 49.4% to 39.3%. Similarly, the 30th percentile of the blood pressure index distribution increases from -0.49 to -0.16

---

[26] The fraction of people with low blood pressure in our sample is negligible; in contrast, 30% are hypertensive (see Figure 8). Therefore we interpret an increase in blood pressure as a worsening of health.

standard deviations. This effect is 2.2 times the average treatment effect (on the treated) of 0.15 standard deviations estimated in Table 2. This result is particularly striking because blood pressure can be controlled through medication, diet, and exercise (Chobanian et al. 2003), and there is a positive association between education and these healthy behaviors (Park and Kang 2008; Conti, Heckman, and Urzua 2010; Cutler and Lleras-Muney 2010).

Figure 6: Distributional Effects on Blood Pressure Index

Notes: The figure shows the pre- and post-reform CDFs of the blood pressure index for compliers. *N* = 270,647.

A comparison of Figures 4 and 6 shows that not only the effect on the body size and blood pressure indices have different signs, but the effects also happen in different parts of the respective distributions. While the effects on body size occur in the upper part of the body size distribution, the effects on blood pressure occur in the lower part of the blood pressure distribution among the healthiest.[27]

---

[27] We do not have the power to test whether some people experienced both a reduction in body size and an increase in blood pressure. However, given that there is a moderate positive correlation between these two outcomes ($\rho = 0.3$), it is unlikely (though possible) that those who experienced reductions in body size were the same individuals experiencing increases in blood pressure. See Appendix Figure D3.

To test whether these shifts in our health indices CDFs are significant and where they are concentrated, we use distributional tests as described in section 3 above. The first row in Table 3 shows p-values of tests of the equality of the pre- and post-reform CDFs. The middle and bottom rows show p-values of tests of the equality of the bottom half (i.e., the healthiest) and the top half (i.e., the least healthy) of pre- and post-reform CDFs.

Table 3: P-values of Distributional Tests

	Body Size	*Lung Function*	*Blood Pressure*
Full Distribution	0.068	0.047	0.036
Bottom Half	0.950	0.509	0.010
Top Half	0.009	0.006	0.148

Notes: The table shows the p-values of tests of the equality of the full distribution, the bottom and top halves of the pre- and post-reform CDFs.

We can reject the null of equality of the full distribution for our three indices (at the 10 % significance level). Consistently with results in Figures 4-6, we can also reject the null for the top halves of the body size and lung function distributions (at the 1% significance level), and for the bottom half of the blood pressure distribution. Figures 7 and 8 shed light on these findings by plotting results for measures with clinical thresholds. Figure 7 shows the pre- and post-reform CDFs of body mass index (for compliers). Figure 8 shows the pre- and post-reform CDFs of diastolic blood pressure (for compliers).[28]

Figure 7 shows that the reductions in BMI caused by more education occur where they matter the most: Staying in school until age 16 *reduces* obesity rates (i.e., the fraction of study participants with a BMI below 30) by 7.5 percentage points. In contrast, Figure 8 shows that the increase in blood pressure does not affect the prevalence of high blood pressure (high blood pressure is classified as having a diastolic blood pressure above 90 mmHg). Staying in school until

---

[28] Results for systolic blood pressure, omitted due to space constraints, are similar.

age 16 *increases* the probability of a "pre-high" blood pressure between 80 mmHg and 90 mmHg by 7.9 percentage points.[29]

**Figure 7: Distributional Effects on Body Mass Index**

These results illustrate the importance of studying distributional effects. While the average treatment effects show improvements in anthropometrics and deterioration in blood pressure, the distributional effects reveal in which part of the distributions these changes occur. The deterioration of blood pressure occurs with no consequences for the prevalence of high blood pressure. In contrast, the improvements in anthropometrics are concentrated at the right tail, with a large reduction in obesity rates. These effects offer important information that policy-makers might wish to trade-off when considering the health consequences of educational policies.

---

[29] Despite being no change in hypertension, the increase in the fraction of people in the pre-high range might still mean a worsening of health, as observational studies indicate that death from both ischemic heart disease (IHD) and stroke increases progressively and linearly from levels as low as 75mmHg DBP (Lewington et al. 2002). In addition, longitudinal data have indicated that DBP between 85 and 89 mmHg are associated with a more than twofold increase in relative risk from cardiovascular disease as compared to those with DBP below 80mmHg (Vasan et al. 2001).

Figure 8: Distributional Effects on Diastolic Blood Pressure
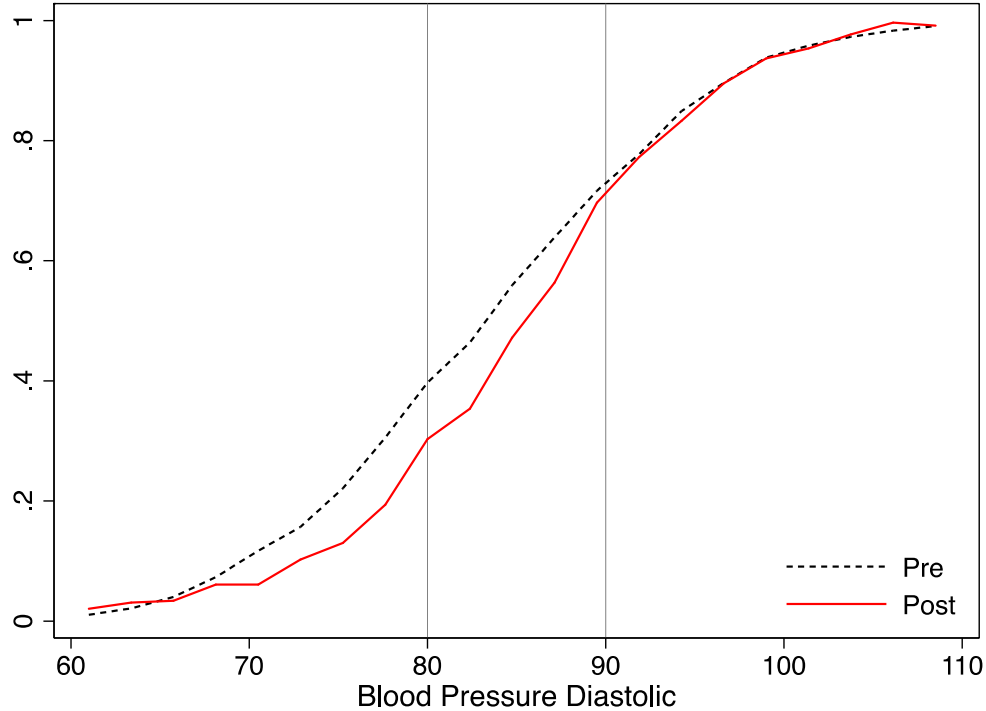
Notes: The figure shows the pre- and post-reform CDFs of diastolic blood pressure for compliers. *N* = 270,647.

5. Channels

In order to better understand the results shown above, in this section we examine possible channels through which education may have affected health. Taking advantage of the richness of the UK Biobank data, we investigate whether education has a causal effect on variables that can be considered proxies for such channels.

We note at the outset the limitations of such exercise. First, there are many channels through which education might affect health and we do not have proxies for all of them. Second, our proxies for the channels are not as good as our health outcome measures – e.g., some of them are self-reported – and some proxies are more credible than others. Third, these channels are inter-related and we are not able to estimate their relative importance for health. Finally, our proxies correspond to a single snapshot in middle age; there may have been differences by education at earlier ages that we cannot measure. We investigate two broadly defined channels for which we have measures: *socioeconomic status* and *health behavior*s.

A. Socioeconomic Status

One channel that has received a lot of attention in the social sciences is that education increases wages and income, giving the more educated greater access to material resources. They can afford more/better quality health care and a healthier diet, for example. They can also afford consuming more "bads" that harm health, such as cigarettes.

Table 4 estimates the effect of staying in school until age 16 on household income.[30] We find that there is an increase in household income, especially at lower income levels. This is not surprisingly, given that the reform increased secondary education only, with no consequences for college attainment. Staying in school until age 16 reduced the likelihoods of living in a household with an annual income below £18,000 and below £31,000 by respectively 7 and 19 percentage points. Estimates of the effect on the likelihoods of living in a household with an annual household income below £52,000 and below £100,000 were small and not statistically significant. These results are broadly consistent with Grenet (2013) that finds that the extra year of schooling induced by the 1972 ROSLA increased wages.

Table 4: Effect on Distribution of Annual Household Income

	Annual household income below			
	£18,000	*£31,000*	*£52,000*	*£100,000*
Stayed in school until 16	-0.070	-0.192	-0.061	0.016
	[0.031]**	[0.043]***	[0.045]	[0.026]
Mean of Y	0.129	0.328	0.622	0.915

Notes: The figure shows the effect of staying in school until age 16 on the distribution of annual household income. $N = 240,880$.

We also find support for the hypothesis that this additional income gives the more educated greater access to material resources. Table 5 shows that staying in school until age 16 decreases the likelihood that the respondent's household has one or fewer cars by 9 percentage points. And even though there is no detectable effect on home ownership[31], staying in school until age 16 may

---

[30] Household income is only reported in 5 brackets: less than £18,000; between £18,000 and £30,999; between £31,000 and £51,999; between £52,000 and £100,000; and more than £100,000.
[31] Home ownership rates are very high in our sample: Around 90% own the place they live in.

have enabled the more educated to live in higher-SES neighborhoods, as measured by the Townsend deprivation index (a higher number corresponds to a greater degree of deprivation).[32] The estimate of this effect is only significant at the 10% level, however. Lastly, there is no evidence that the more educated live in neighborhoods with less pollution.[33]

Table 5: Effect on Car and Home Ownership, Neighborhood SES and Pollution

	Number of Cars				*Home Ownership*	*Townsend*	*Pollution Index*
	= 0	*≤ 1*	*≤ 2*	*≤ 3*			
Stayed in school until 16	-0.030	-0.091	-0.017	-0.017	0.005	-0.497	-0.056
	[0.022]	[0.041]**	[0.033]	[0.019]	[0.026]	[0.254]*	[0.088]
Mean of Y	0.0790	0.412	0.801	0.947	0.899	-1.331	5.48e-11
N Observations		270,055			269,363	270,705	248,333

Notes: The figure shows estimates of the staying in school until age 16 on car and home ownership and neighborhood SES and pollution.

Education may also enable the more educated to get "better jobs" that offer safer work environments and where they interact with more educated co-workers, which in itself may have positive effects on one's health behaviors. A job with higher SES may also improve one's relative position in society, which could improve health through a reduction in hierarchy-related stress (Marmot et al. 1978, 1991).

To investigate the hypothesis that the more educated work in better jobs, we classify the socioeconomic status of the occupations in which study participants worked using the 2000 National Statistics Socio-economic Classification (NS-SEC), the primary social classification in the United Kingdom.[34] Table 6 estimates the effect of staying in school until age 16 on the SES of

---

[32] The Townsend Deprivation Index is a measure of the material deprivation of the neighborhoods in which study participants lived. It is constructed from four rates measured at the neighborhood level: (1) unemployment, (2) non-car ownership, (3) non-home ownership, and (4) household overcrowding rates. These rates were estimated for each output area using data from the 2001 national census. Participants were assigned the scores of the output areas where their residential postcodes were located.

[33] The pollution index is a summary index of nitrogen oxides air pollution (UKB field 24004), particulate matter with diameter less than or equal to 2.5 micrometres (UKB field 24006), particulate matter with diameter less than or equal to 10 micrometres (UKB fields 24005 and 24019), and nitrogen dioxide air pollution (UKB fields 24003, 24016, 24017, and 24018). See http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=114 for more details. Averages were taken whenever there were measures for multiple years.

[34] Respondents who were employed or self-employed were asked in a verbal interview to describe their jobs. Interviewers were provided with an algorithm to code respondent's answers according to a tree structure, categorized following the Standard Occupation Classification 2000. We construct a measure of socio-economic

occupations.[35] Lower values correspond to higher SES – see Appendix Table D2 for specific categories.

Table 6: Effect on Occupation SES

	Socioeconomic Class					
	= 7	≥ 6	≥ 5	≥ 4	≥ 3	≥ 2
Stayed in school until 16	-0.032	-0.073	-0.084	-0.117	-0.070	-0.041
	[0.023]	[0.037]**	[0.040]**	[0.043]***	[0.050]	[0.044]
Mean of Y	0.0545	0.163	0.203	0.251	0.437	0.750

Notes: The figure shows estimates of the staying in school until age 16 on the socioeconomic class of the participants' occupations. $N = 207,533$.

Staying in school until age 16 increases the likelihood that the individual works in higher-SES occupations. The estimates are uniformly negative, implying that an additional year of education reduces the likelihood of working in a lower-SES occupation and surpass significance at the 5% level at three out of the 6 SES thresholds considered. The effect is concentrated at the bottom part of the distribution in semi-routine occupations, lower supervisory, and technical occupations.[36]

B. Health Behaviors

Education may also change health behaviors. Correlational evidence shows that the more educated are more likely to use preventive care, that they manage chronic conditions more effectively, and that they are less likely to smoke and drink heavily (Cutler and Lleras-Muney 2008; Goldman and Smith 2002). Education may lead to healthy behaviors by providing people with more knowledge and better critical thinking skills to absorb information. Education may also change time and risk preferences (Becker and Mulligan 1994; Perez-Arce 2017) with consequences for health investments. Another potential channel is peer effects, which can affect health behaviors (Jensen and Lleras-Muney 2012). By working in higher-ranked occupations and

---

class from these job codes, by matching 300+ codes into 7 SEC classes (The National Statistics Socio-Economic Classification User Manual, 2005) that have a hierarchical structure from lower to higher SES.

[35] One issue in studying the effects of education on occupations is that education may also affect the likelihood that one is working. We find no effects of education on employment or retirement (results available upon request).

[36] However, we do not find an effect of education on the (self-reported) likelihood that as part of his/her job a worker had to stand/walk or do heavy physical activity (results available upon request).

living in better neighborhoods, those affected by the reform might have healthier peers, which might generate and/or strengthen improvements in health behaviors.[37]

We study the effects of education on three types of health behaviors: diet, smoking, and physical activity. Diet was measured using a 24-h dietary assessment tool self-completed through the Internet (Galante et al. 2016).[38] Accelerometers worn for 7 days were used to measure physical activity.[39] Smoking was self-reported. Given the estimated increase in blood pressure, we also examine effects on hypertension diagnosis by a doctor and blood pressure medication (both self-reported).

Table 7: Effect on Diet

	Calories	% Sugars	% Fat	% Saturated Fat	% Carbo-hydrates
Stayed in school until 16	-86.868 [153.979]	0.020 [0.017]	-0.030 [0.017]*	-0.019 [0.008]**	0.011 [0.021]
Mean of Y	2108	0.221	0.329	0.126	0.483

Notes: The figure shows estimates of the effects of staying in school until age 16 on diet. Study participants were asked about their diet in five different waves (at baseline and four online surveys), such that there are sometimes multiple observations by participant. For this reason, standard errors are clustered at the individual level. $N = 268,957$ observations, corresponding to 122,665 study participants.

Table 7 shows the effects on diet. Staying in school until age 16 reduces the intake of fat and saturated fat (as a fraction of total energy intake) by 3 and 2 percentage points respectively. There are no effects on caloric intake, sugars, or carbohydrates. Similarly, Table 8 shows that we can detect no effects on smoking, physical activity, and hypertension diagnosis and medication given the measures that we have available. These are consistent with the results in Clark and Royer (2013) that show no effects of the 1972 ROSLA on self-reported smoking or physical activity.

---

[37] Other channels through which education might affect health are marriage and fertility. Geruso and Royer (2017) show that the 1972 ROSLA increased spousal education, lowered teen fertility rates but had no impact on marriage rates or completed fertility by age 45.

[38] The Oxford WebQ collects information on the quantities of all foods and beverages consumed over the previous day. Respondents are asked whether they consumed any of 21 food groups over the previous day. A positive response results in the screen expanding to reveal a list of commonly consumed foods in the corresponding category. Respondents then select the amount of each food consumed using standard categories to indicate the amount consumed. Energy and nutrient values are generated by multiplying the quantity of each food or drink consumed by its nutrient composition. The Oxford WebQ was included at the assessment visit of the baseline measures for the last 70,724 participants and administered over the Internet to all UK Biobanks participants with a known email address, who were invited to complete the Oxford WebQ on four separate occasions over a 16-month period.

[39] Accelerometer data were collected from May 2013 until December 2015 from 103,720 UK Biobank participants. Our outcome of interest is the average acceleration adjusted for no-wear bias (UKB field 90087): http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=90087

Table 8: Effect on Smoking, Physical Activity, and Hypertension Diagnosis and Medication

	Currently Smoke	Ever Smoke	Hypertension Diagnosis	Hypertension Medication	Physical Activity
Stayed in school until 16	-0.002	0.043	-0.035	0.002	0.456
	[0.027]	[0.042]	[0.033]	[0.022]	[2.704]
Mean of Y	0.118	0.396	0.205	0.0724	28.87
N Observations	270,937	267,384	270,700	268,315	61,701

Notes: The figure shows estimates of the effects of staying in school until age 16 on smoking, physical activity, and hypertension diagnosis and medication.

The analysis suggests that improved diet is an important channel through which education reduces body size. Our estimates show that those who stayed in school longer due to the reform had better diets in middle age – about 10% lower in fat and 15% lower in saturated fat (see Table 7). Even if the energy content of one's diet is held constant, changes in diet composition can affect body weight (Hall et al. 2012).

The reduced body size could, in turn, improve lung function. Weight loss has been shown to be associated with improved lung function (Thomas et al. 1989). Alternatively, the improved lung function could reflect reductions in smoking that we cannot capture with our self-reported measures. Smoking has been shown to be widely under-reported (Jackson and Beaglehole 1985), and this under-reporting could vary by education. Since spirometry is strongly affected by smoking, our objective measures of lung function might be better at capturing changes in behavior than self-reported measures of smoking are.

The pathways are less clear for the harmful effect of education on blood pressure. Blood pressure is different from body size in the sense that it can be more easily controlled through medical treatment. Consistent with our finding of no difference in clinically measured hypertension in middle age, we find no increase in a (self-reported) measure of hypertension diagnosis.[40] Moreover, we find no effect on (current) blood pressure medication. We conclude that our finding that education raises blood pressure cannot be explained by differences in hypertension diagnosis and treatment by education.

---

[40] The wording of the question was "Have you been told by a doctor that you have high blood pressure"?

One alternative hypothesis is that, by changing the types of occupations and careers individuals have, education might have an effect on job responsibilities, expectations, and work-related stress with negative implications for blood pressure.[41] If the additional schooling increases the likelihood of working in occupations with hierarchies above, it could increase rank-related stress (Marmot et al. 1978, 1991). In the U.S. context, for example, academically successful African Americans have higher biomarkers related to cardiometabolic risk (i.e. blood pressure and stress hormones) than other groups (Brody et al. 2013, Miller et al. 2015, Chen et al. 2015). This is potentially driven by stressors related to upward mobility, which could also be playing a role in the U.K context. We have no credible data to test this hypothesis in the UK Biobank so we leave it for future work.

6. Conclusions

In this paper, we investigate how education affects the distribution of health along three dimensions: body size, lung function, and blood pressure. We find that an increase in secondary education has important implications for health in middle age. Our results offer three takeaways that add new insights to our understanding about the relationship between education and health.

First, *it is important to consider different health dimensions separately, as education might improve some dimensions while worsening others*. Our findings show that the additional schooling generated by the 1972 ROSLA improved body size and lung function, while increasing blood pressure. These findings emerge from our mean analysis as well as our distributional analysis. While education seems to have reduced body size through improvements in SES and diet, it is not clear the channels through which education increased blood pressure. Replicating such finding in other contexts and exploring the channels through which education can harm some health dimensions while improving others is an important avenue for future research.

Second, *the focus of the current literature on mean effects can lead to misleading conclusions.* In particular, the effects of education are concentrated in parts of the health distribution: The effects on these parts can be as much as 2.5 times larger than the average treatment effect, meaning that even in samples as large as those used here, average effects are only robustly detected at the 10% confidence-level. Additionally, even though we cannot estimate

---

[41] Table 6 presents evidence that education increases the socioeconomic class of occupations participants hold in middle age. They are less likely to work on semi-routine and lower supervisory occupations, for example.

individual-level effects, our results *suggest* that there is substantial variation in the effects across people. While the effects may be large for some, others may not be affected at all.

Third, *the distributional effects of education on health might be different for different health dimensions*. We present evidence that while the improvement in body size and lung function are concentrated at the upper tail of their respective distributions (among the least healthy), the worsening in blood pressure is concentrated at the lower tail of the blood pressure distribution (among the most healthy).

Overall, our results suggest that – *despite the increase in blood pressure* – schooling may be an effective policy tool to improve health. Because the reductions in BMI occur where they matter the most, staying in school until age 16 reduces obesity rates by 7.5 percentage points. It also improves lung function. In contrast, the increase in blood pressure is concentrated below the clinical threshold for hypertension (i.e., diastolic blood pressure above 90 mmHg or systolic blood pressure above 140) with no consequences for hypertension prevalence.

REFERENCES

Allen et al. 2012: UK Biobank: Current status and what it means for epidemiology

Albouy, Valerie and Laurent Lequien. 2009. "Does compulsory education lower mortality?" *Journal of Health Economics*, 28,155–168.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." Journal of the American statistical Association 103.484, 1481-1495.

Anderson, Theodore W., and Donald A. Darling. 1952. "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes." The annals of mathematical statistics. 193-212.

Anderson, T.W.; Darling, D.A. 1954. "A Test of Goodness-of-Fit". Journal of the American Statistical Association. 49: 765–769. doi:10.2307/2281537.

Becker, G.S., Mulligan, C.B. 1997. "The endogenous determination of time preference." *Quarterly Journal of Economics* 112 (August (3)), 729–758.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review*, 96(4): 988-1012.

Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). Experimental evidence on distributional effects of Head Start (No. w20434). National Bureau of Economic Research.

BLACK, D.A., J.A. SMITH, M.C. BERGER AND B. J. NOEL (2003): "Is the threat of reemployment services more effective than the services themselves? Experimental evidence from the UI system," *American Economic Review* 93(3), 1313–1327.

Brody, G. H., Yu, T., Chen, E., Miller, G. E., Kogan, S. M., & Beach, S. R. (2013). Is resilience only skin deep?: Rural African Americans' socioeconomic status-related risk and competence in preadolescence and psychological adjustment and allostatic load at age 19. *Psychological Science, 24,* 1285–1293. http://dx.doi.org/10.1177/0956797612471954

Chen, E., Miller, G. E., Brody, G. H., & Lei, M. K. (2015). Neighborhood poverty, college attendance, and diverging profiles of substance use and allostatic load in rural African American youth. *Clinical Psychological Science, 3,* 675–685.http://dx.doi.org/10.1177/2167702614546639

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo Jr, J. L., ... and Roccella, E. J. 2003. "The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the JNC 7 report." **Jama**, 289(19), 2560-2571.

Clark, Damon and Heather Royer. 2010. "The Effect of Education on Adult Health and Mortality: Evidence from Britain." NBER Working Paper No. 16013.

Clark, Damon and Heather Royer. 2013. "The Effect of Education on Adult Mortality and Health: Evidence from Britain." American Economic Review, 103(6): 2087–2120.

Conti, G., Heckman, J., and Urzua, S. 2010. "The education-health gradient." *The American economic review*, 100(2), 234.

Conti, G., Heckman, J. J., & Urzua, S. (2011). Early endowments, education, and health. Unpublished manuscript, University of Chicago, Department of Economics.

Cutler, David M., Adriana Lleras-Muney, and Tom Vogl. 2008. Socioeconomic Status and Health: Dimensions and Mechanisms. Working Paper 14333. http://www.nber.org/papers/w14333

Cutler, David M., and Adriana Lleras-Muney. 2010. "Understanding differences in health behaviors by education." *J Health Econ*. 29(1):1-28. PMCID: PMC2824018

Cutler, David M., and Adriana Lleras-Muney. 2010. "The education gradient in old age disability." *Research Findings in the Economics of Aging.* Chicago: University of Chicago Press, 101-120. http://www.nber.org/chapters/c8194

Davies, Neil M., Matt Dickson, George Davey Smith, Gerard van den Berg, and Frank Windmeijer. 2016. "The Causal Effects of Education on Health, Mortality, Cognition, Well-being, and Income in the UK Biobank." bioRxiv doi: https://doi.org/10.1101/074815

De Matteis, S., Jarvis, D., Hutchings, S., Darnton, A., Fishwick, D., Sadhra, S., ... & Cullinan, P. (2016). Occupations associated with COPD risk in the large population-based UK Biobank cohort study. *Occupational and environmental medicine*, oemed-2015.

Dickson, M. and Smith, S., 2011. What determines the return to education: An extra year or a hurdle cleared?. *Economics of education review*, *30*(6), pp.1167-1176.

DJEBBARI, H. AND J.A. SMITH (2008): "Heterogeneous impacts in PROGRESA," *Journal of Econometrics* 145(1), 64–80.

FIRPO, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Galante, J., Adamska, L., Young, A., Young, H., Littlejohns, T. J., Gallacher, J., & Allen, N. (2016). The acceptability of repeat Internet-based hybrid diet assessment of previous 24-h

dietary intake: administration of the Oxford WebQ in UK Biobank. *British Journal of Nutrition*, 115(04), 681-686.

Gelman, Andrew and Guido Imbens. 2016. "Why high-order polynomials should not be used in regression discontinuity designs." Unpublished Manuscript.

Geruso, Michael and Heather Royer. 2017. "The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK." Unpublished Manuscript.

Grenet, Julien. 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.

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

Grossman, Michael. 2015. "The Relationship between Health and Schooling: What's New?" No. w21609. National Bureau of Economic Research. http://www.nber.org/papers/w21609

Hall, Kevin D., Steven B. Heymsfield, Joseph W. Kemnitz, Samuel Klein, Dale A. Schoeller, and
John R. Speakman. 2012. "Energy Balance and its Components: Implications for Body Weight Regulation—Consensus Statement from the American Society for Nutrition." *American Journal of Clinical Nutrition*, 95, 989–994.

HECKMAN, J.J., J. SMITH AND N. CLEMENTS (1997): "Making the most out of programme evaluations and social exper-iments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies* 64(4), 487–535.

JACKSON, R., & BEAGLEHOLE, R. (1985). Secular trends in underreporting of cigarette consumption. *American journal of epidemiology*, *122*(2), 341-344.

Jensen, R., & Lleras-Muney, A. (2012). Does staying in school (and not working) prevent teen smoking and drinking?. *Journal of Health Economics*, 31(4), 644-657.

Johnston, David W., Carol Propper, and Michael A. Shields. 2009. "Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient." *Journal of health economics* 28.3, 540-552.

Jürges, Hendrik, Eberhard Kruk and Steffen Reinhold. 2013. "The effect of compulsory schooling on health—evidence from biomarkers." *J Popul Econ*, 26:645–672.

Kemptner, Daniel, Hendrik Jürgesb and Steffen Reinholdc. 2011. "Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany." *Journal of Health Economics*, 30, 340–354.

Jensen, R. and Lleras-Muney, A., 2012. "Does staying in school (and not working) prevent teen smoking and drinking?" *Journal of Health Economics*, 31(4): 644-657.

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142 675 – 697. The regression discontinuity design: Theory and applications

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

Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. Prospective Studies Collaboration. *Lancet* 2002;360:1903-13.

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

Marmot M.G., G. Rose, M. Shipley, and P.J.S. Hamilton. 1978. "Employment grade and coronary heart disease in British civil servants." *Journal of Epidemiology and Community Health* 1978 32: 244-249.

Marmot, M.G., et al. 1991. "Health inequalities among British civil servants: The Whitehall II study." *Lancet* 337(8754): 1387-1393.

Miller, G. E., Yu, T., Chen, E., & Brody, G. H. (2015). Self control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 10325–10330. http://dx.doi.org/10.1073/ pnas.1505063112

Office of National Statistics. 2005. The National Statistics Socio-Economic Classification User Manual.http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/ guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec- -rebased-on-soc2010--user-manual/index.html

Park, C., and Kang, C. 2008. "Does education induce healthy lifestyle?" *Journal of Health Economics*, 27(6), 1516-1531.

Perez-Arce, Francisco. 2017. "The effect of education on time preferences." *Economics of Education Review.* Vol. 56 (2), 52-64.

Powdthavee, Nattavudh. 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. http://www.jstor.org/stable/10.1086/657020

Shen, Shu, and Xiaohan Zhang. 2016. "Distributional tests for regression discontinuity: Theory and empirical examples." Review of Economics and Statistics 98.4: 685-700.

Silles, Mary A. 2009. "The causal effect of education on health: Evidence from the United Kingdom." *Economics of Education Review*, 28(1):122–128.

Stephens, Michael A. "EDF statistics for goodness of fit and some comparisons." *Journal of the American statistical Association* 69.347 (1974): 730-737.

Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, Rory 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. PMCID:PMC4380465

Thomas, P. S., Cowen, E. R., Hulands, G., & Milledge, J. S. (1989). Respiratory function in the morbidly obese before and after weight loss. *Thorax*, *44*(5), 382-386.

Vasan RS, Larson MG, Leip EP, Evans JC, O'Donnell CJ, Kannel WB, et al. Impact of highnormal blood pressure on the risk of cardiovascular disease. *N Engl J Med* 2001;345:1291-7.

Appendix A

Appendix Figure A1: McCrary Test
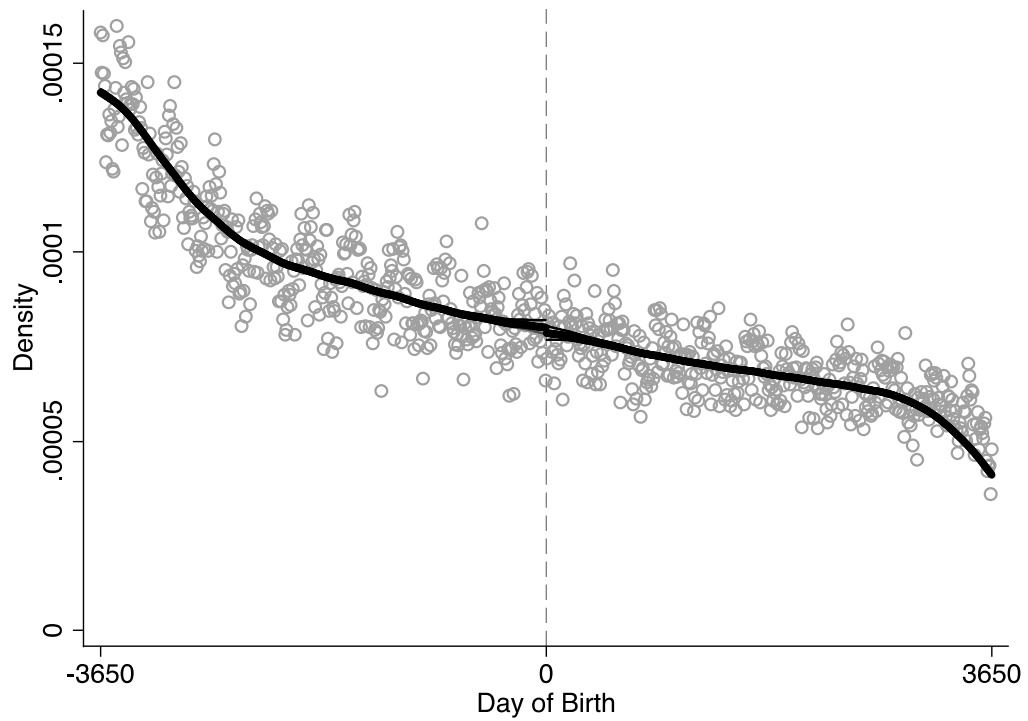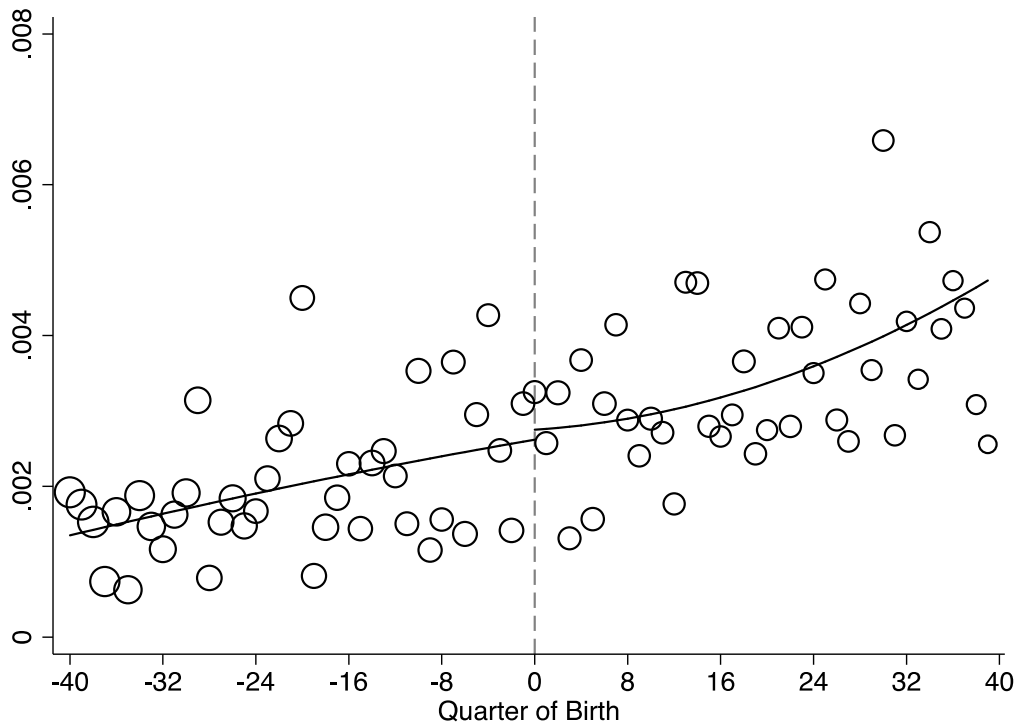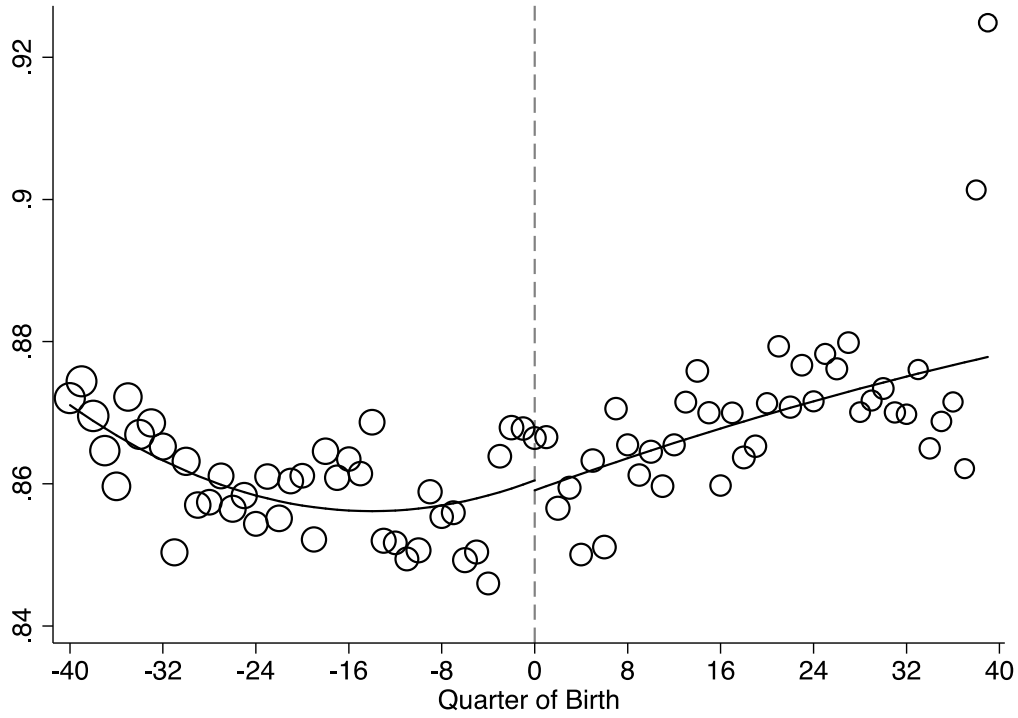

Notes: The figure shows the fraction of study participants by day of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show nonparametric birth cohort trends. The estimated discontinuity of the density is -0.0201 with a standard error of 0.0174. $N = 271,234$.

Appendix Figure A2: Male
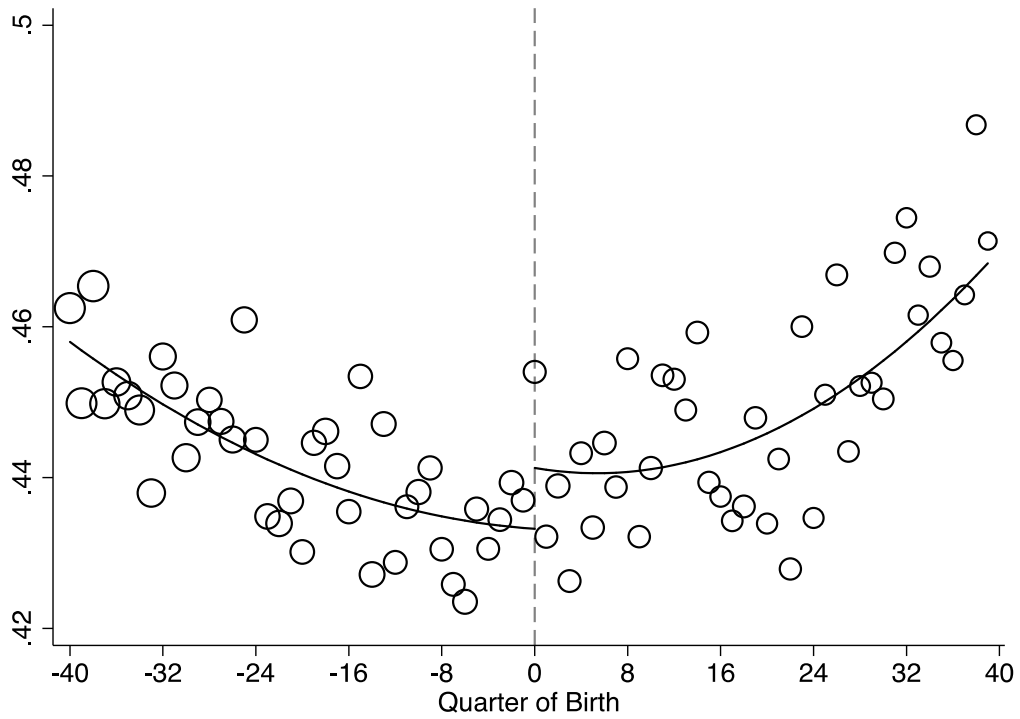
Notes:
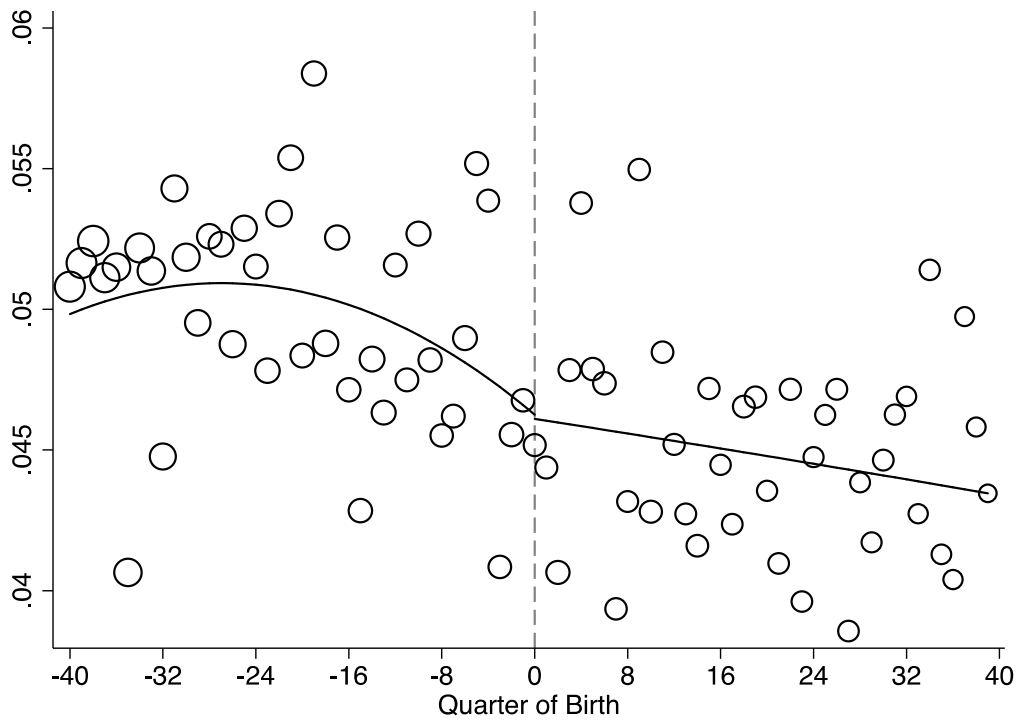*Notes*: The figure shows the fraction of male study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
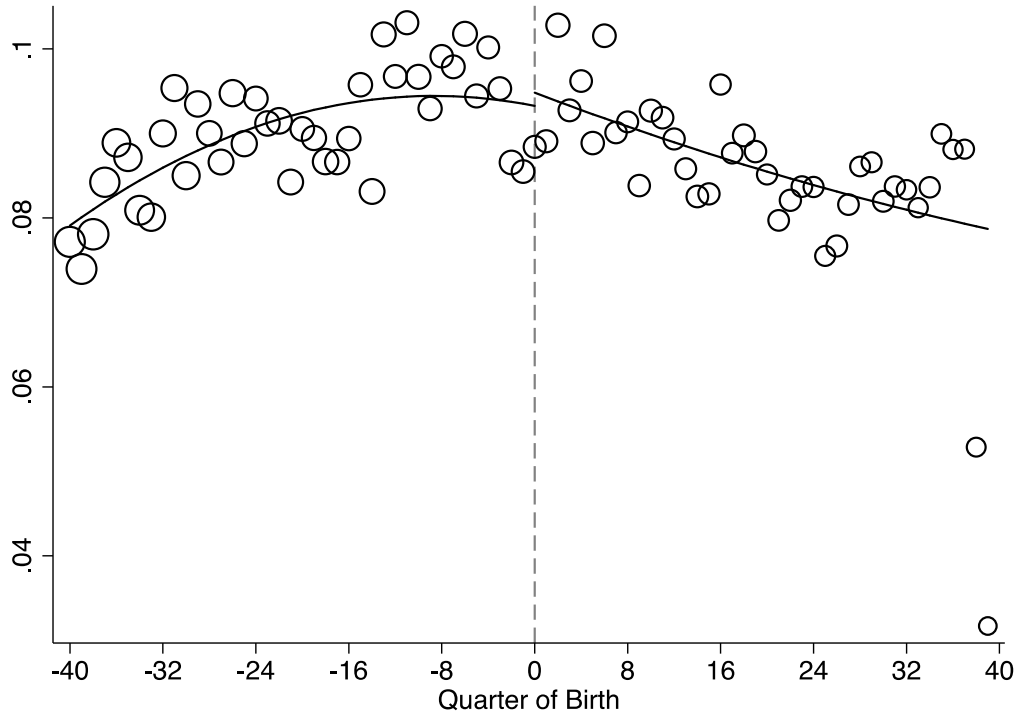
Appendix Figure A3: White
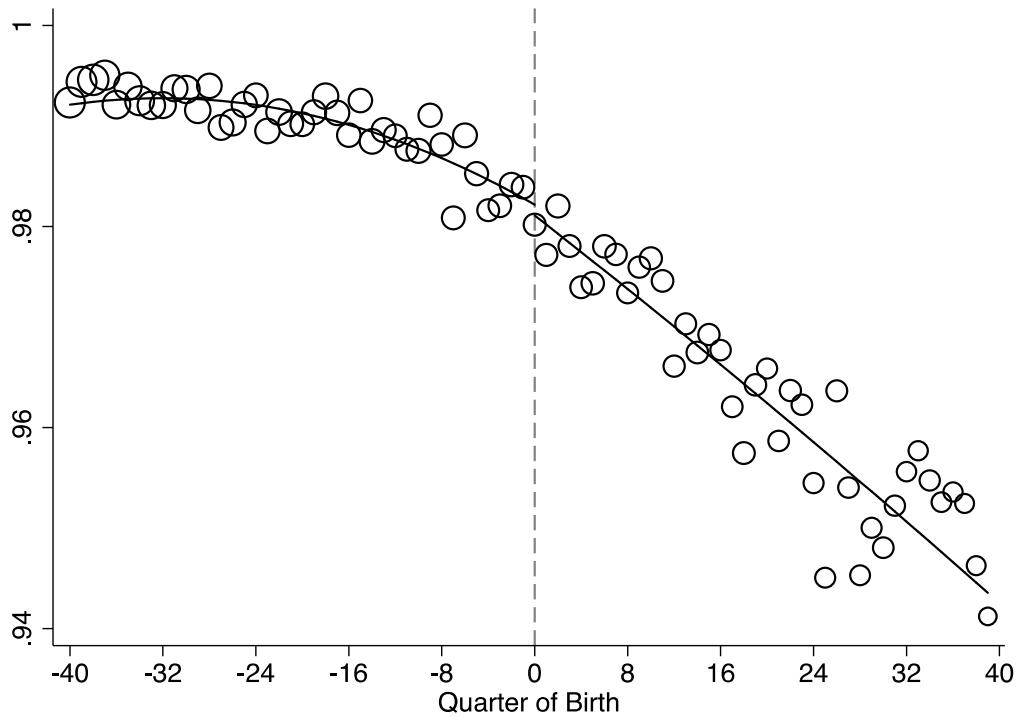
Notes: The figure shows the fraction of white study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure A4: Mixed Ethnicity
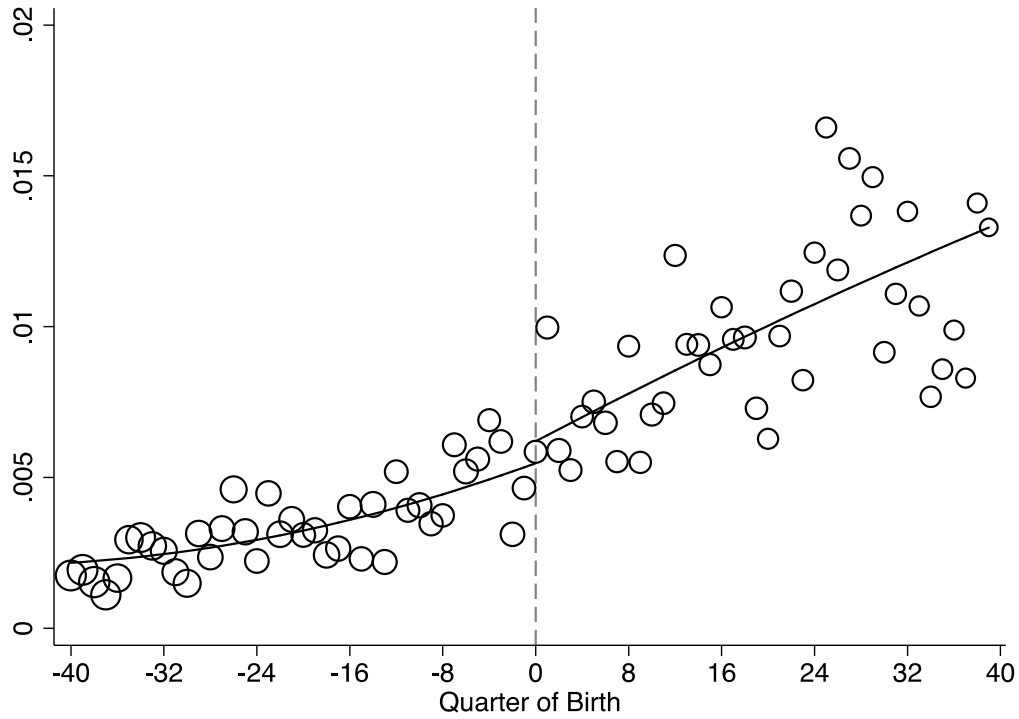
Notes: The figure shows the fraction of study participants of mixed ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
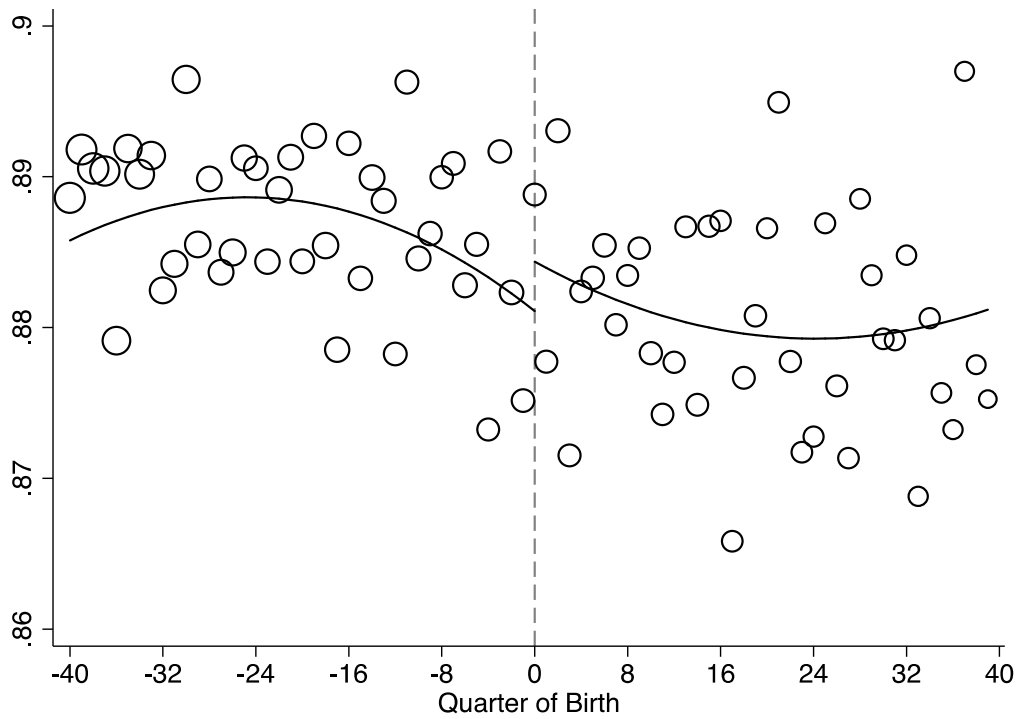
Appendix Figure A5: Asian
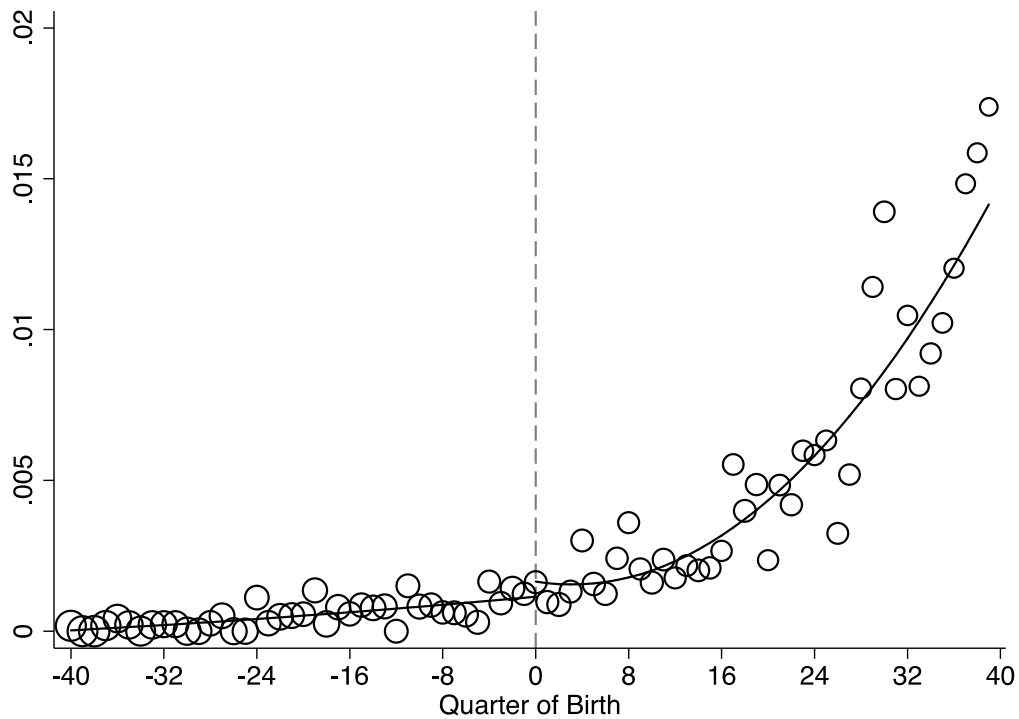
Notes: The figure shows the fraction of Asian study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
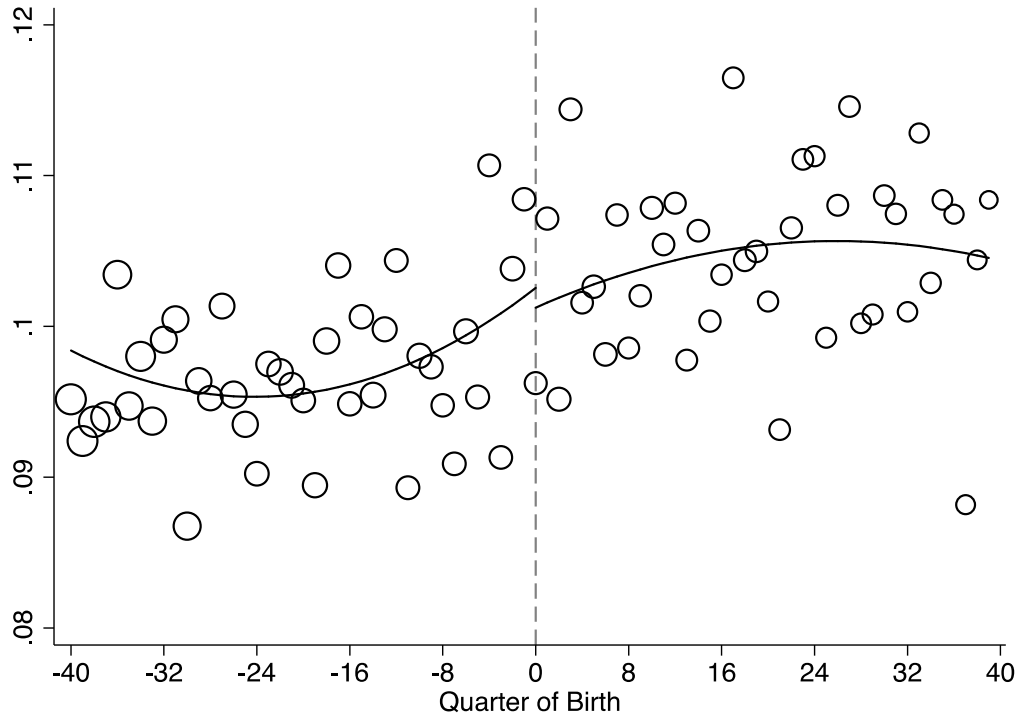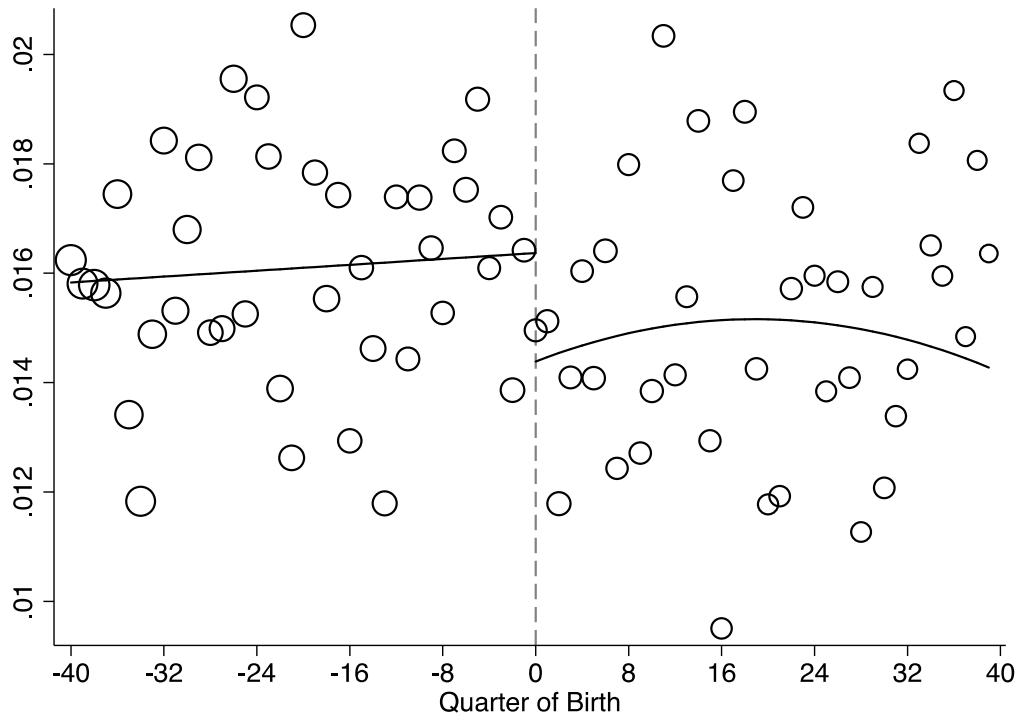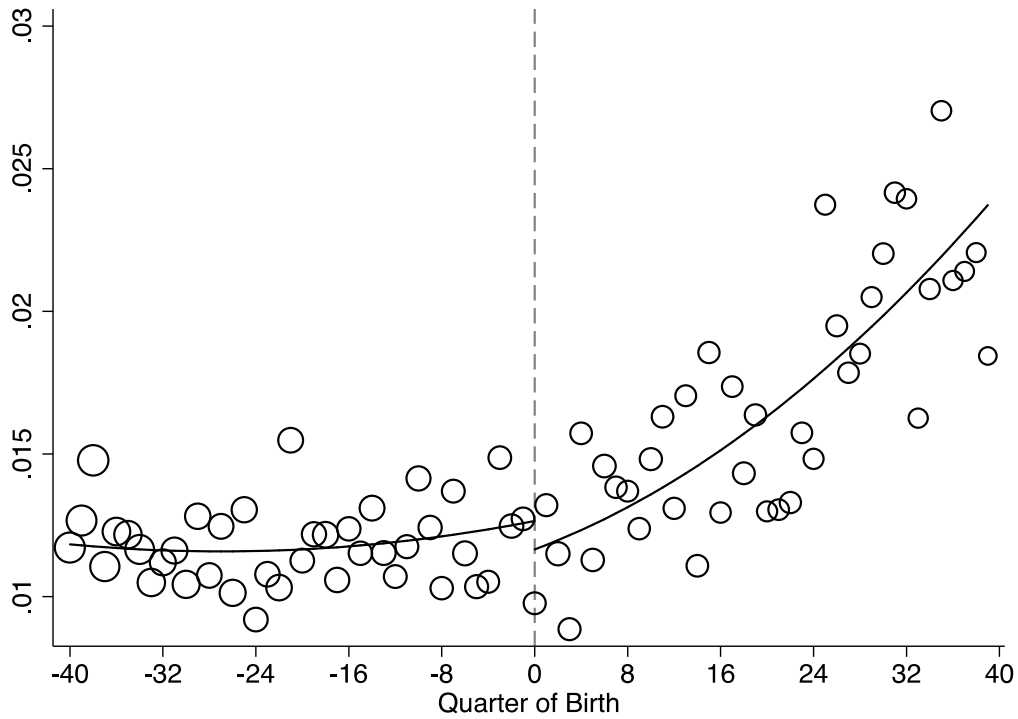
Appendix Figure A6: Black
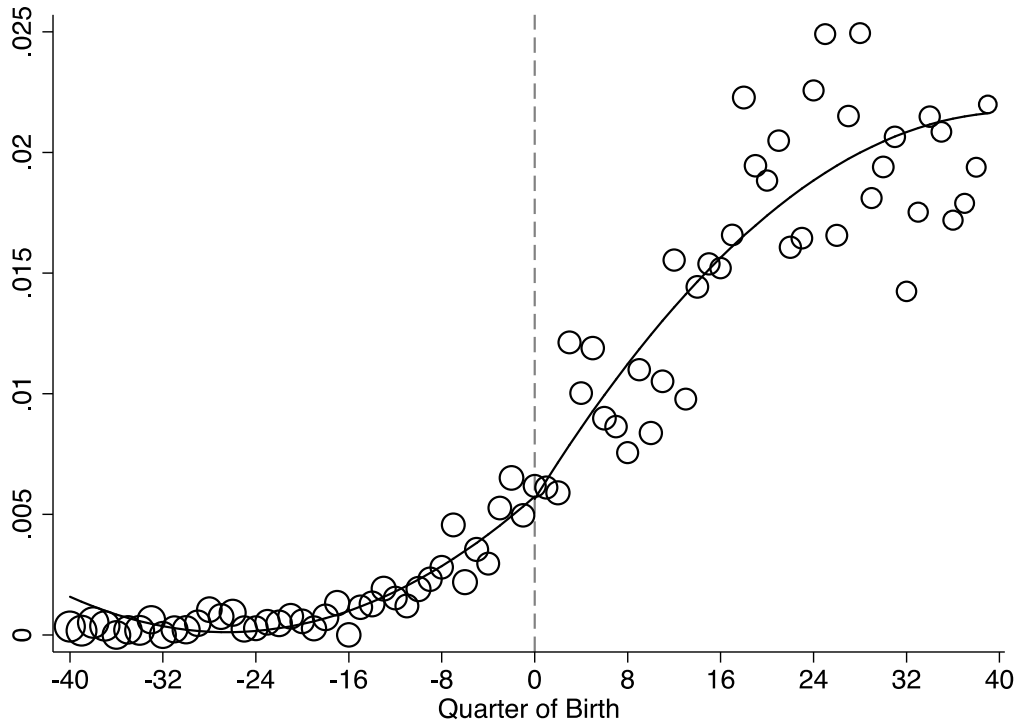
Notes: The figure shows the fraction of black study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure A7: Other Ethnicity

Notes: The figure shows the fraction of study participants of another ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure A8: Born in England

Notes: The figure shows the fraction of study participants born in England by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
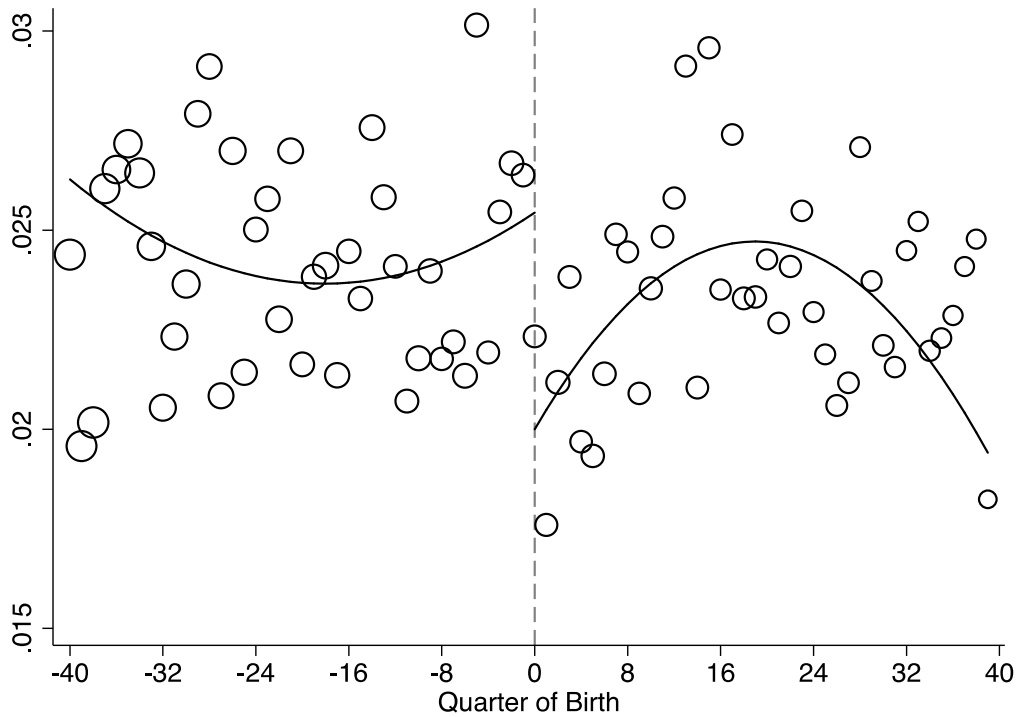
Appendix Figure A9: Born in Wales

Notes: The figure shows the fraction of study participants born in Wales by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure A10: Born in Scotland

Notes: The figure shows the fraction of study participants born in Scotland by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271{,}082$.

Appendix Figure A11: Right Handed

Notes: The figure shows the fraction of right-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271{,}023$.

Appendix Figure A12: Left Handed

Notes: The figure shows the fraction of left-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,023$.

Appendix Figure A13: Ambidextrous

Notes: The figure shows the fraction of ambidextrous study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,023$.

Appendix Figure A14: Adopted

Notes: The figure shows the fraction of study participants who were adopted by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 270,723$.

Appendix Figure A15: Twin

Notes: The figure shows the fraction of study participants who were twins by quarter of birth. This question was not asked to those who had been adopted. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 267,130$.

Appendix Table A1: Balance Control Test

	Male	White	Mixed Ethnicity	Asian	Black	Other Ethnicity	Born in England
Post	0.008	-0.001	0.001	0.001	-0.000	0.000	-0.001
	[0.006]	[0.002]	[0.001]	[0.000]	[0.001]	[0.001]	[0.004]
N	271,082	271,082	271,082	271,082	271,082	271,082	271,082
Mean of Y	0.436	0.983	0.00511	0.00130	0.00501	0.00274	0.862

	Born in Wales	Born in Scotland	Right Handed	Left Handed	Ambi-dextrous	Adopted	Twin
Post	-0.000	0.001	0.003	-0.001	-0.002	-0.001	-0.006
	[0.003]	[0.004]	[0.004]	[0.004]	[0.002]	[0.001]	[0.002]***
N	271,082	271,082	271,023	271,023	271,023	270,723	267,130
Mean of Y	0.0466	0.0915	0.881	0.103	0.0158	0.0127	0.0252

Notes: The table investigates whether predetermined characteristics are smooth are around the September 1, 1957 cutoff. It reports the coefficient on an indicator for being born on or after September 1, 1957 (i.e., "Post") from regressions where the dependent variables is listed in the column. The regressions also included quadratic polynomials in date of birth, which were allowed to differ on either side of the cutoff. The mean of Y corresponds to the average of the dependent variable among those born in the 12 months before September 1, 1957.

Appendix Figure A16: East Coordinate of Birth Place
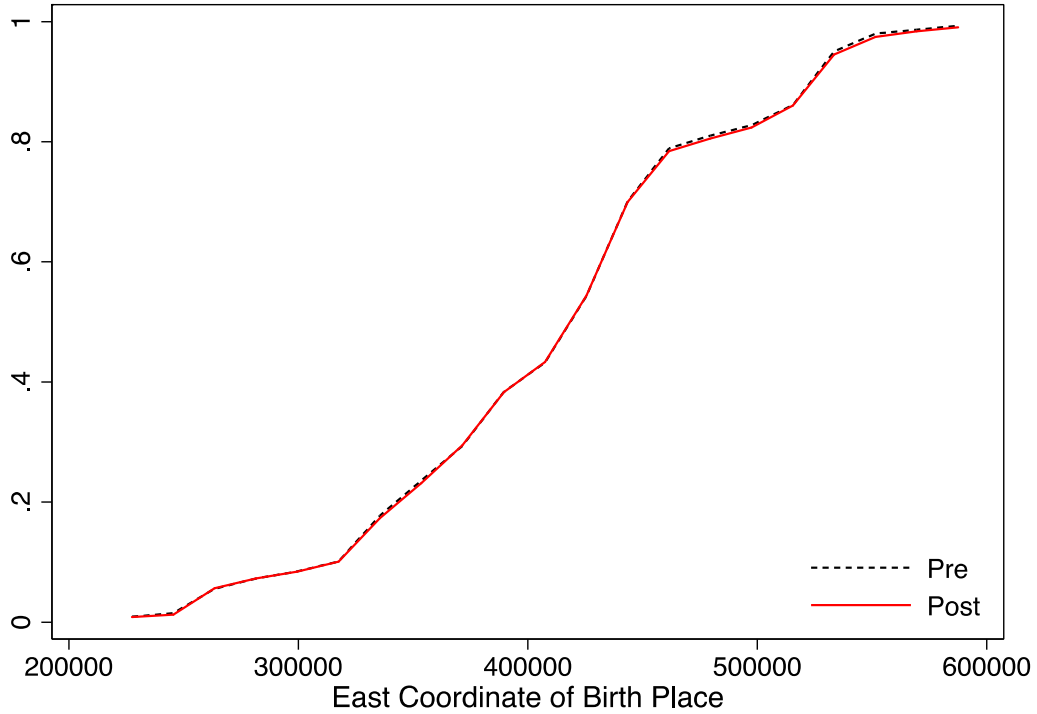
Notes: The figure shows the pre- and post-reform CDFs of east coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 266,883$.
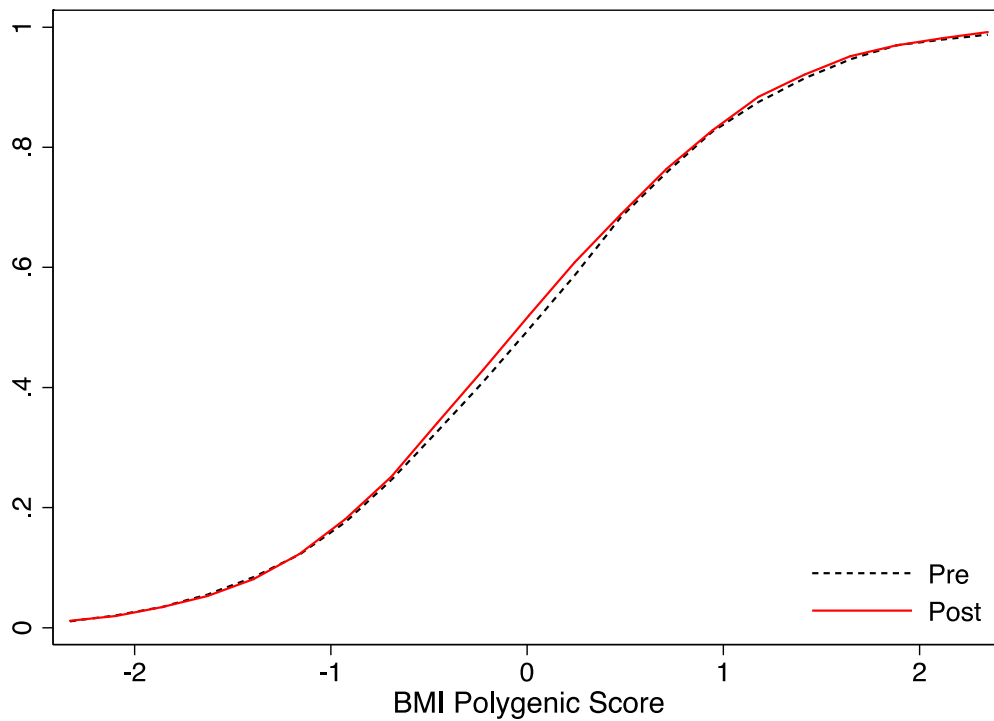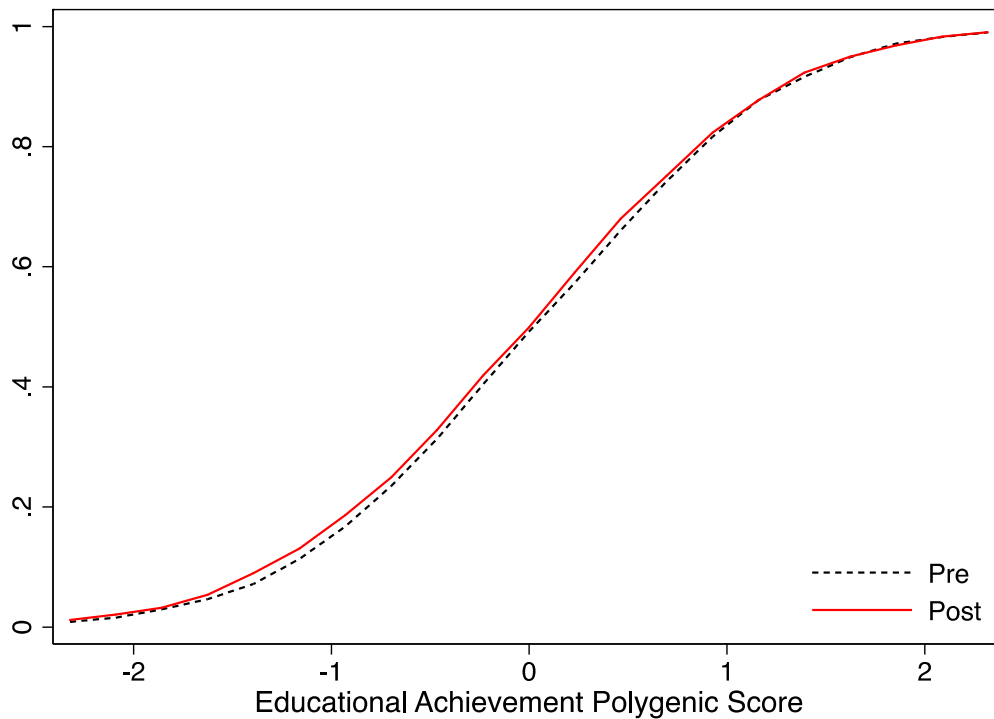
Appendix Figure A17: North Coordinate of Birth Place
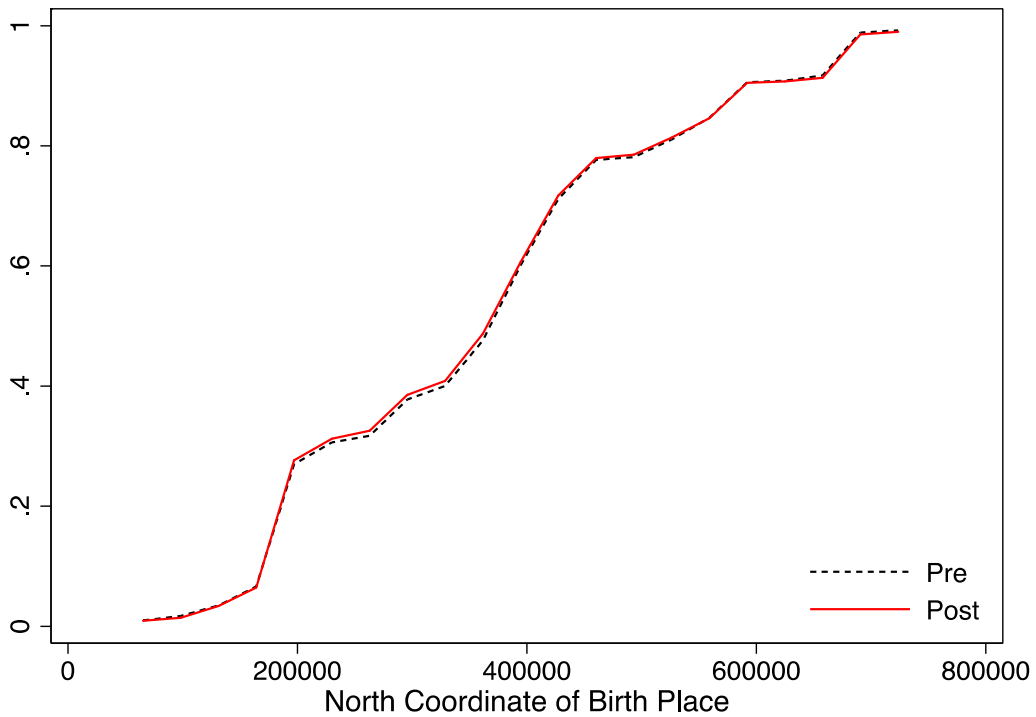
Notes: The figure shows the pre- and post-reform CDFs of north coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 266,883$.

Appendix Figure A18: Subischial Height
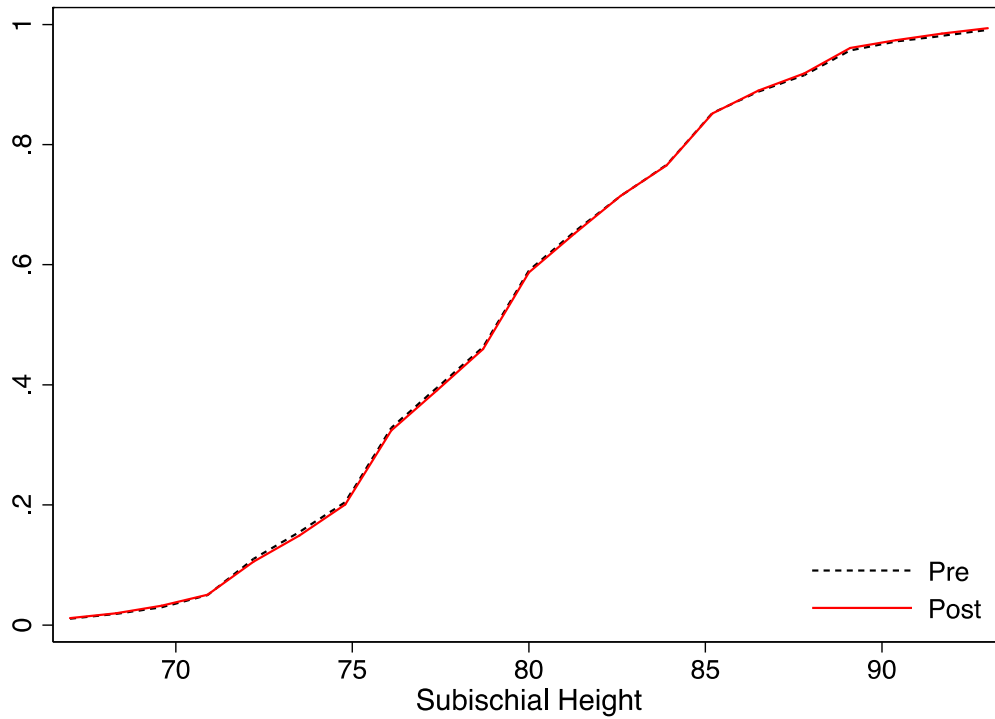
Notes: The figure shows the pre- and post-reform CDFs of subischial height. Subischial height is the difference between standing height and sitting height. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. *N* = 269,173.

Appendix Figure A19: Fraction Missing Genetic Data
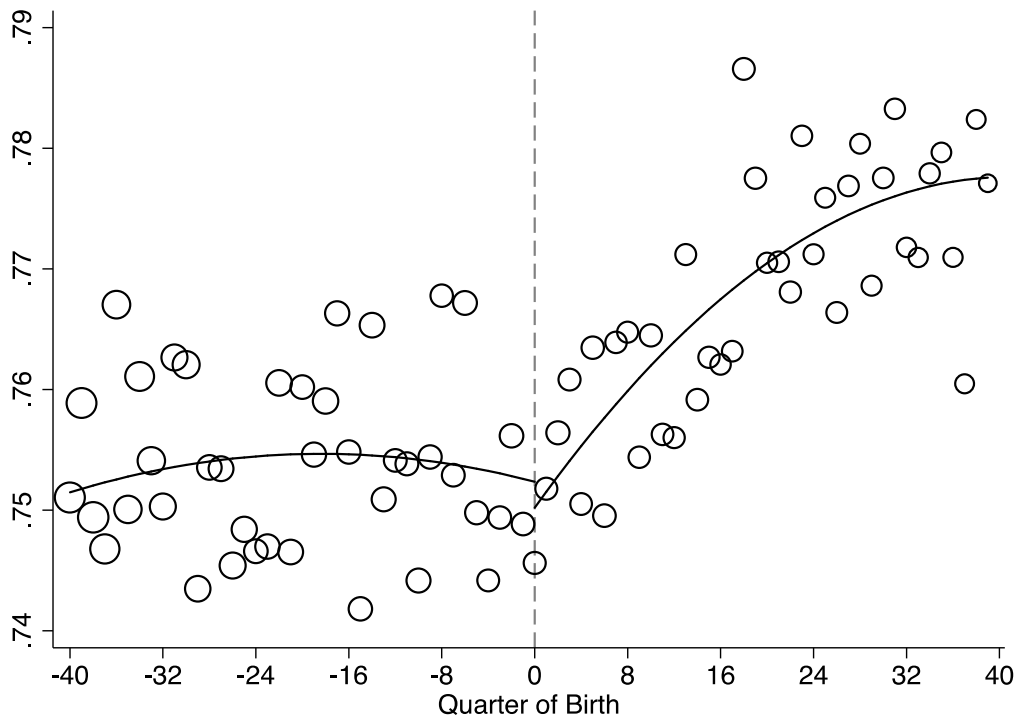
Notes: The figure shows the fraction of study participants with genetic data available by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The

circumference of each circle reflects the number of participants born in that quarter. The discontinuity is -0.0028 with a standard error of 0.0054 (p-value of 0.61). The mean among those born in the 12 months before the cutoff is 0.7539. $N = 271,244$.

Appendix Figure A20: Body Mass Index Polygenic Score

Notes: The figure shows the pre- and post-reform CDFs of the polygenic score for BMI. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 65,138$. Notice that the genetic data was available for only ¼ of the sample.

Appendix Figure A21: Educational Achievement Polygenic Score

Notes: The figure shows the pre- and post-reform CDFs of the polygenic score for educational achievement. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 65,138$. Notice that the genetic data was available for only ¼ of the sample.

Appendix Table A2: Distributional Test

Coordinates of Birth Place		Subischial	Polygenic Scores	
East	*North*	*Height*	*BMI*	*Educational Achievement*
0.23	0.65	0.51	0.31	0.19

Notes: The table show the p-values of tests of the equality of the pre- and post-reform CDFs. N = 266,883 (coordinates of place of birth); 269,173 (subischial height); and 65,138 (polygenic scores for BMI and educational achievement)

Appendix B

Appendix Figure B1: Average of Body Size Index by Quarter of Birth

Notes: The figure assesses the sensititivity of the results for body size index to the choice of bandwidth and to the use of linear trends. It shows the average of body size index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.
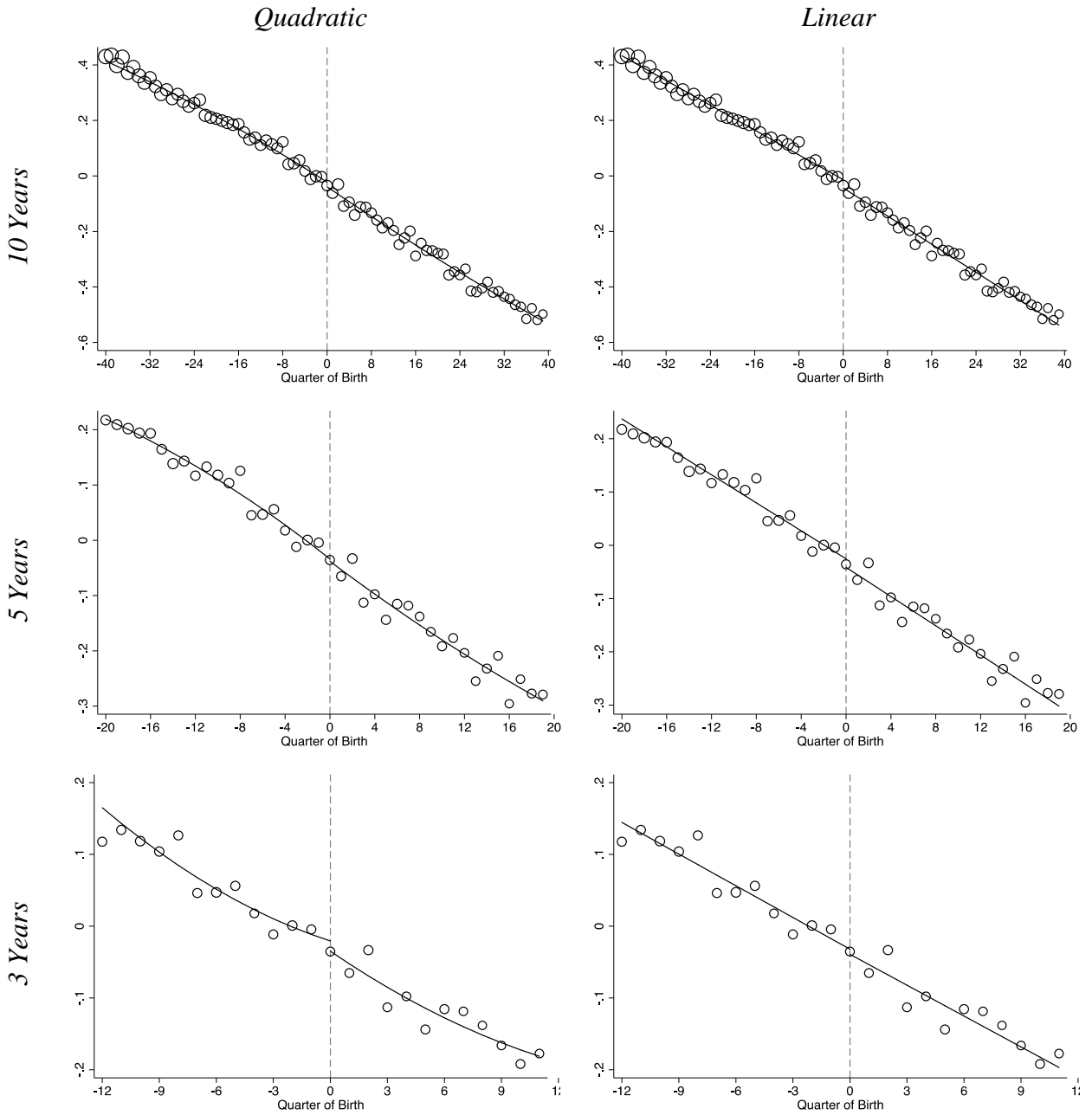
Appendix Figure B2: Average of Lung Function Index by Quarter of Birth
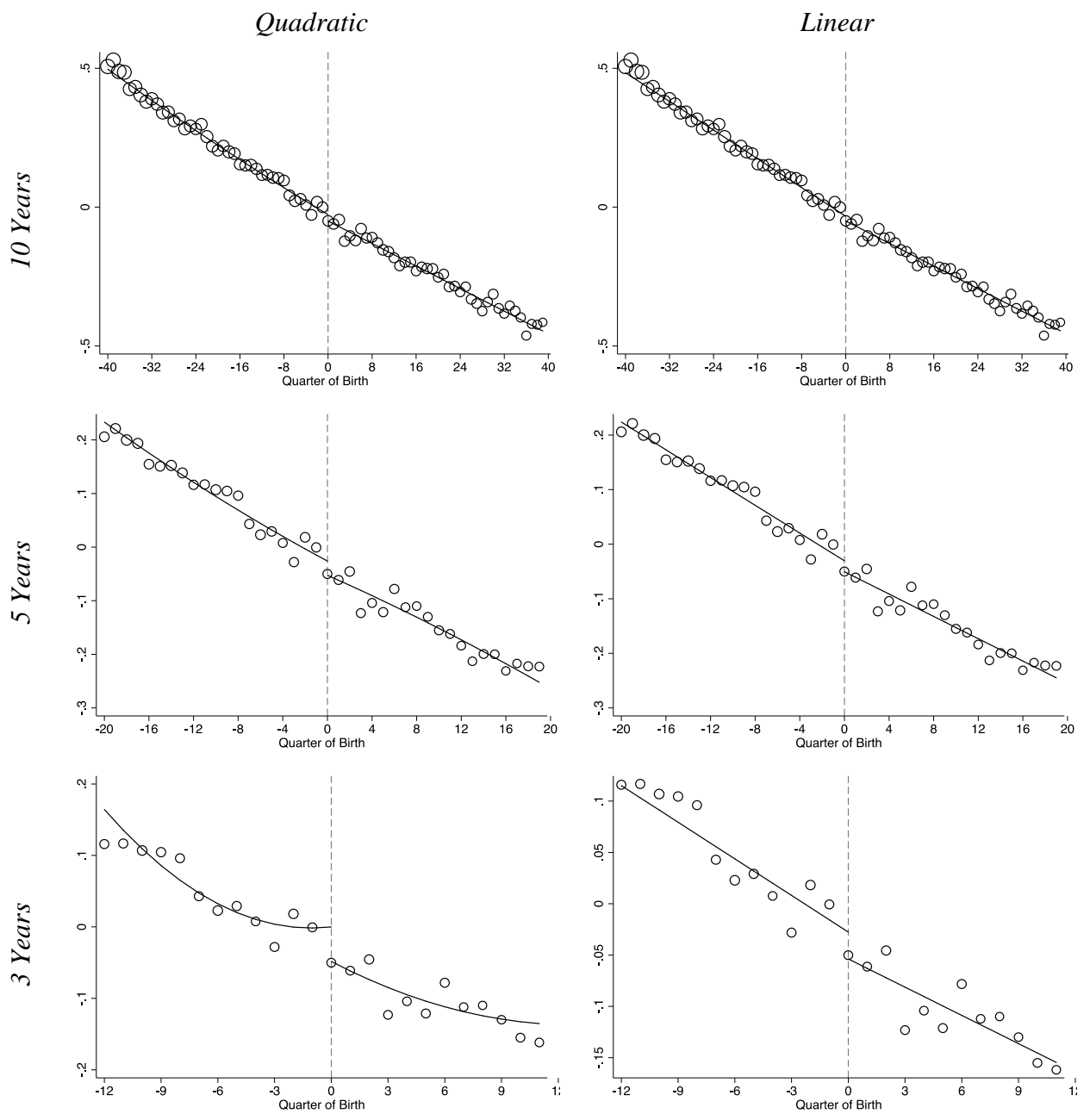
Notes: The figure assesses the sensititivity of the results for lung function index to the choice of bandwidth and to the use of linear trends. It shows the average of lung function index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

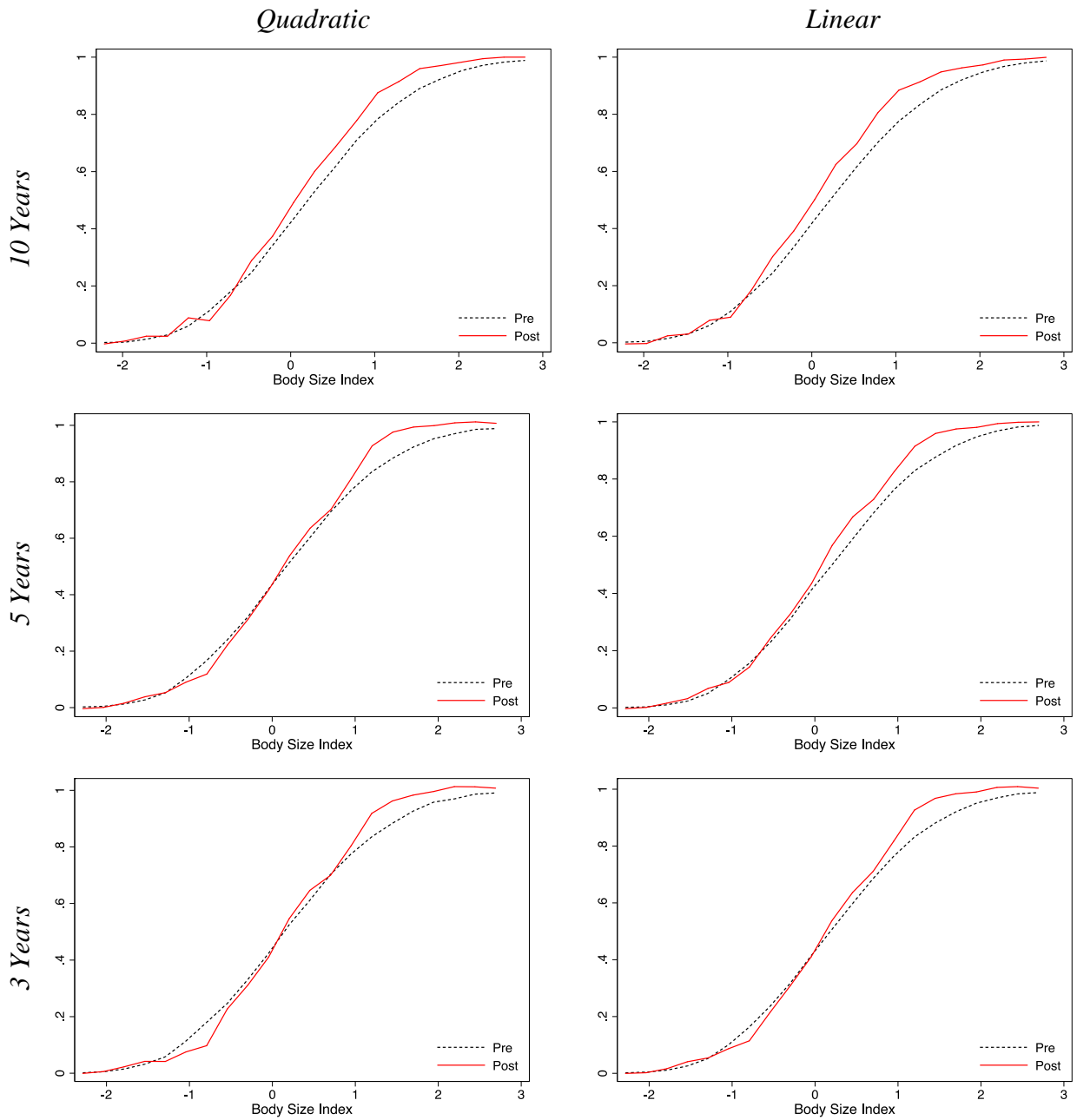# Appendix Figure B3: Average of Blood Pressure Index by Quarter of Birth

Notes: The figure assesses the sensititivity of the results for blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the average of blood pressure index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

Appendix Figure B4: Average of Summary Index by Quarter of Birth

Notes: The figure assesses the sensititivity of the results for the summary index to the choice of bandwidth and to the use of linear trends. It shows the average of the summary index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

Appendix Figure B5: Distributional Effects on Body Size (No Controls)
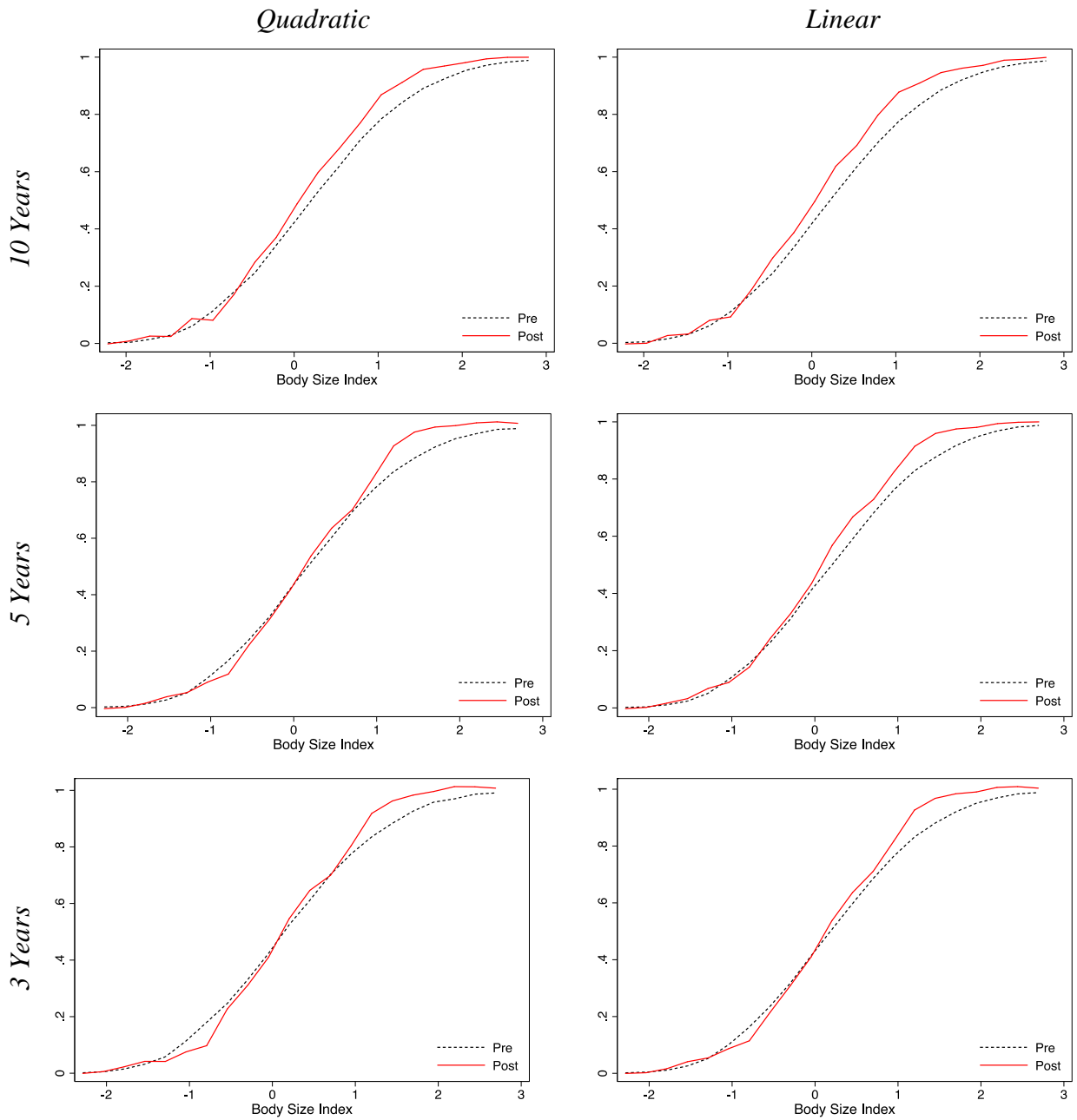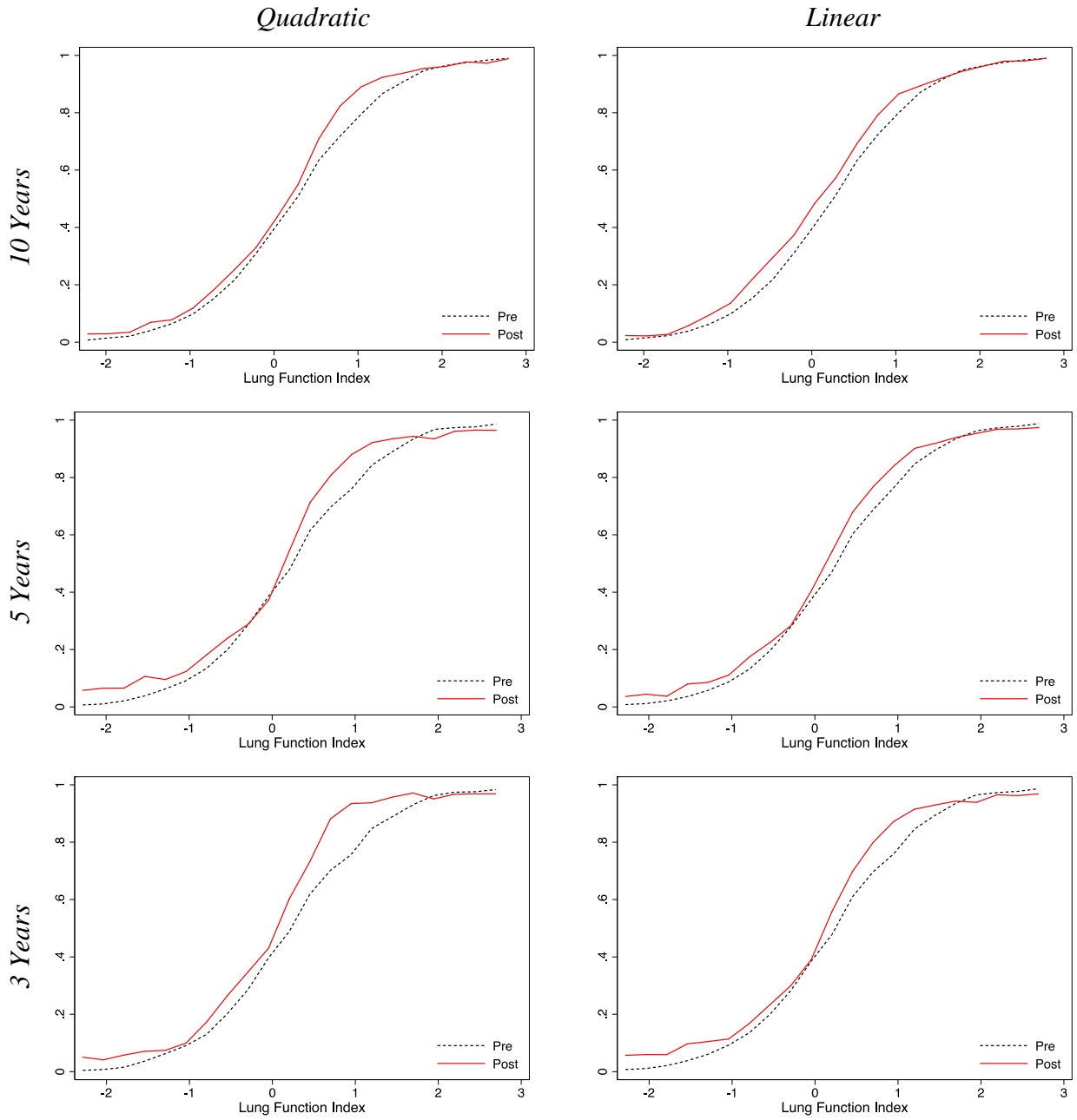

Notes: The figure assessess the sensititivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. No controls.

Appendix Figure B6: Distributional Effects on Body Size (With Controls)

Notes: The figure assesses the sensititivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The regressions include the following set of controls: gender, age in days (at the time of the baseline assessment) and age squared, dummies for ethnicity, dummies for country of birth, and dummies for calendar month of birth.

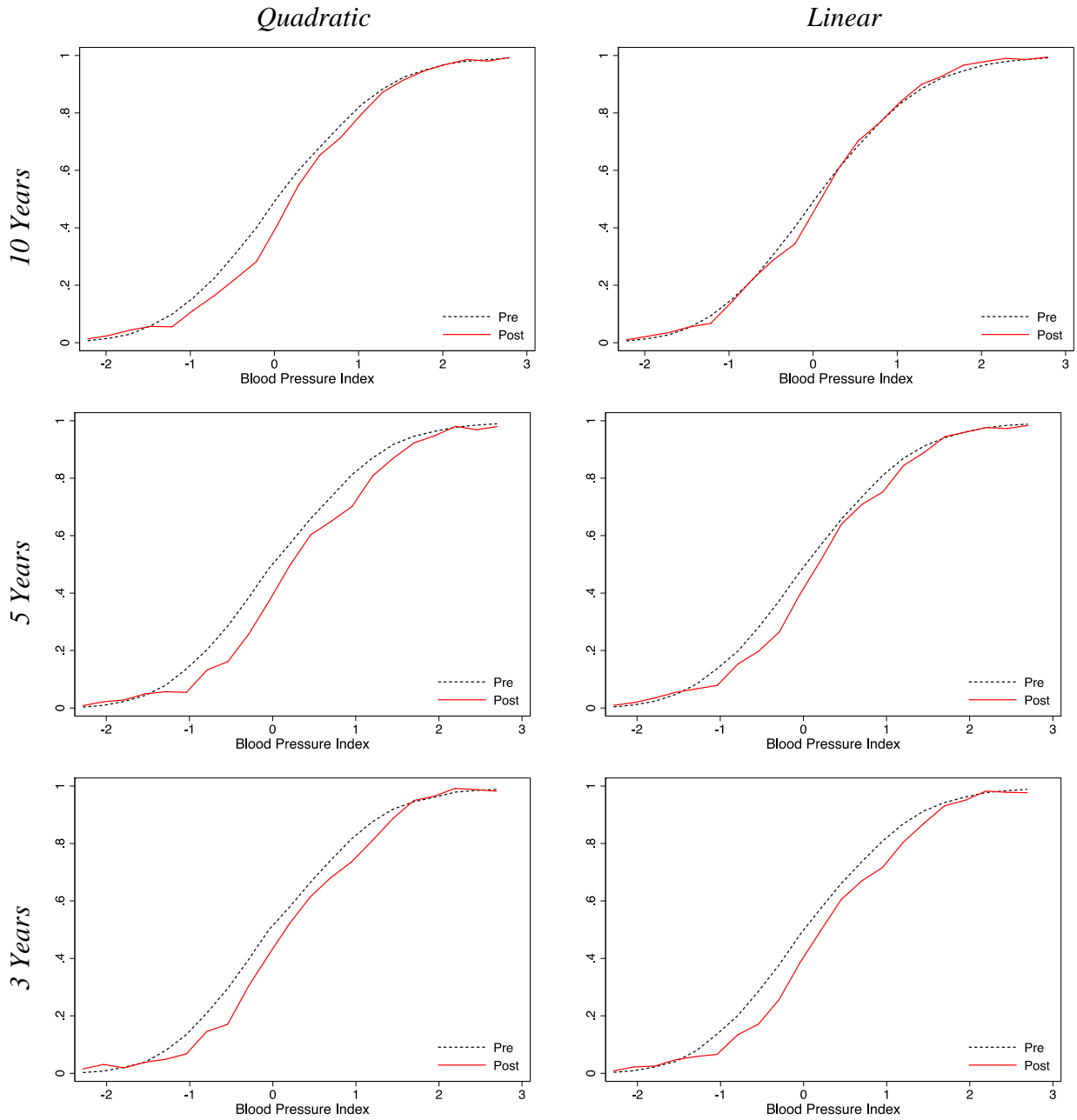# Appendix Figure B7: Distributional Effects on Lung Function (No Controls)
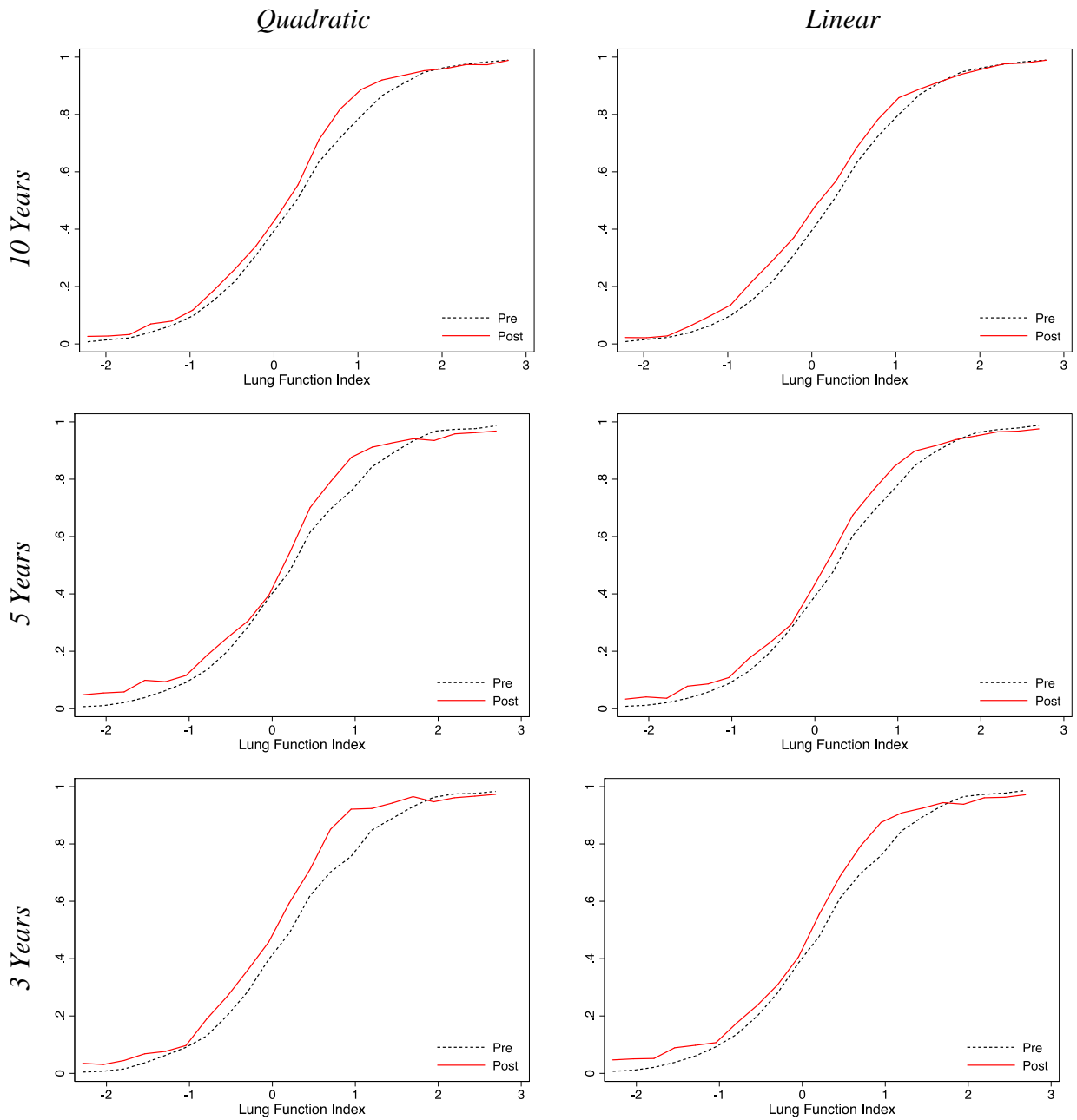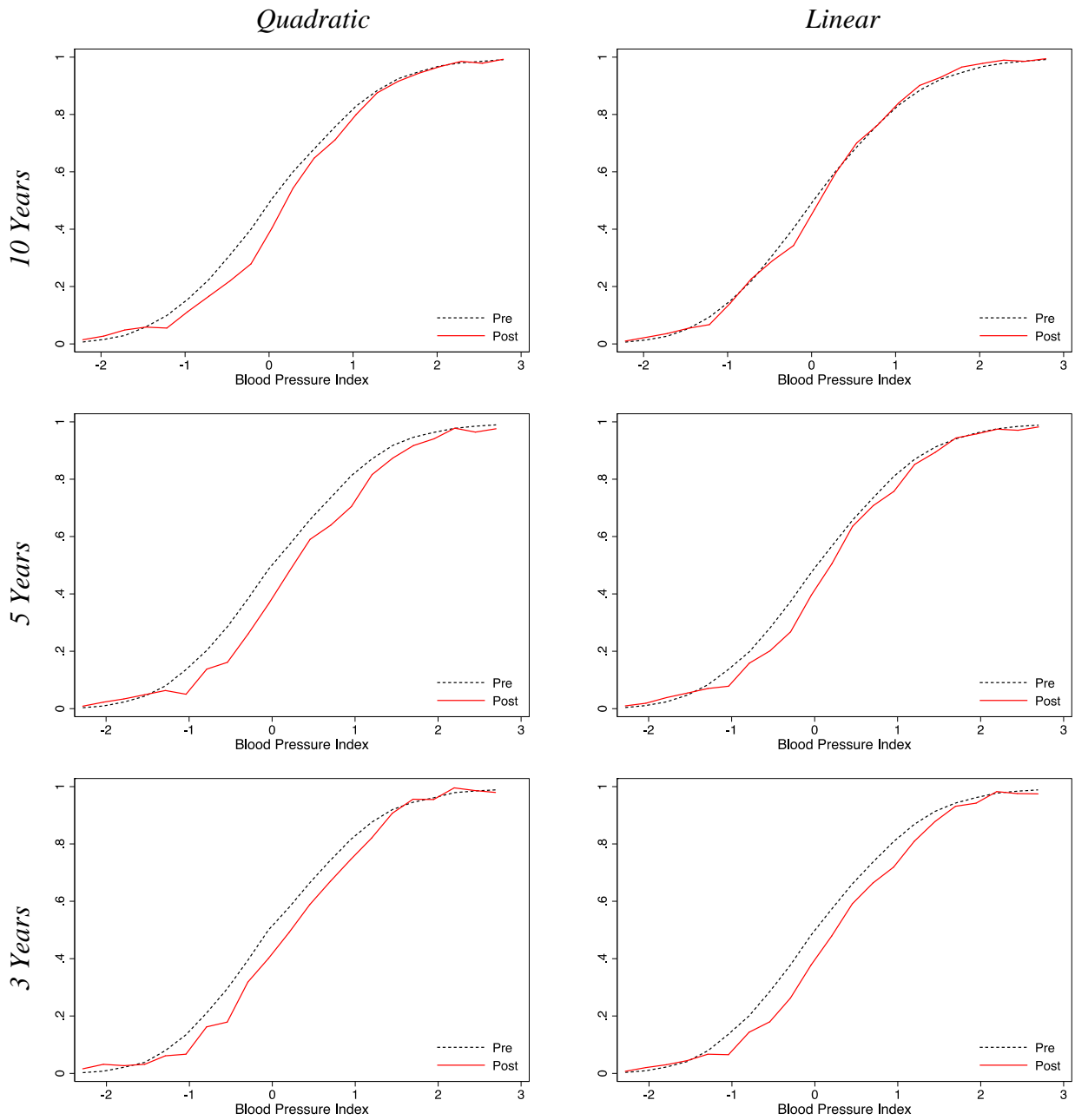

Notes: The figure assesses the sensititivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the lung function index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. No controls.

Appendix Figure B8: Distributional Effects on Lung Function (With Controls)

Notes: The figure assesses the sensititivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the lung function index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The regressions include the following set of controls: gender, age in days (at the time of the baseline assessment) and age squared, dummies for ethnicity, dummies for country of birth, and dummies for calendar month of birth.

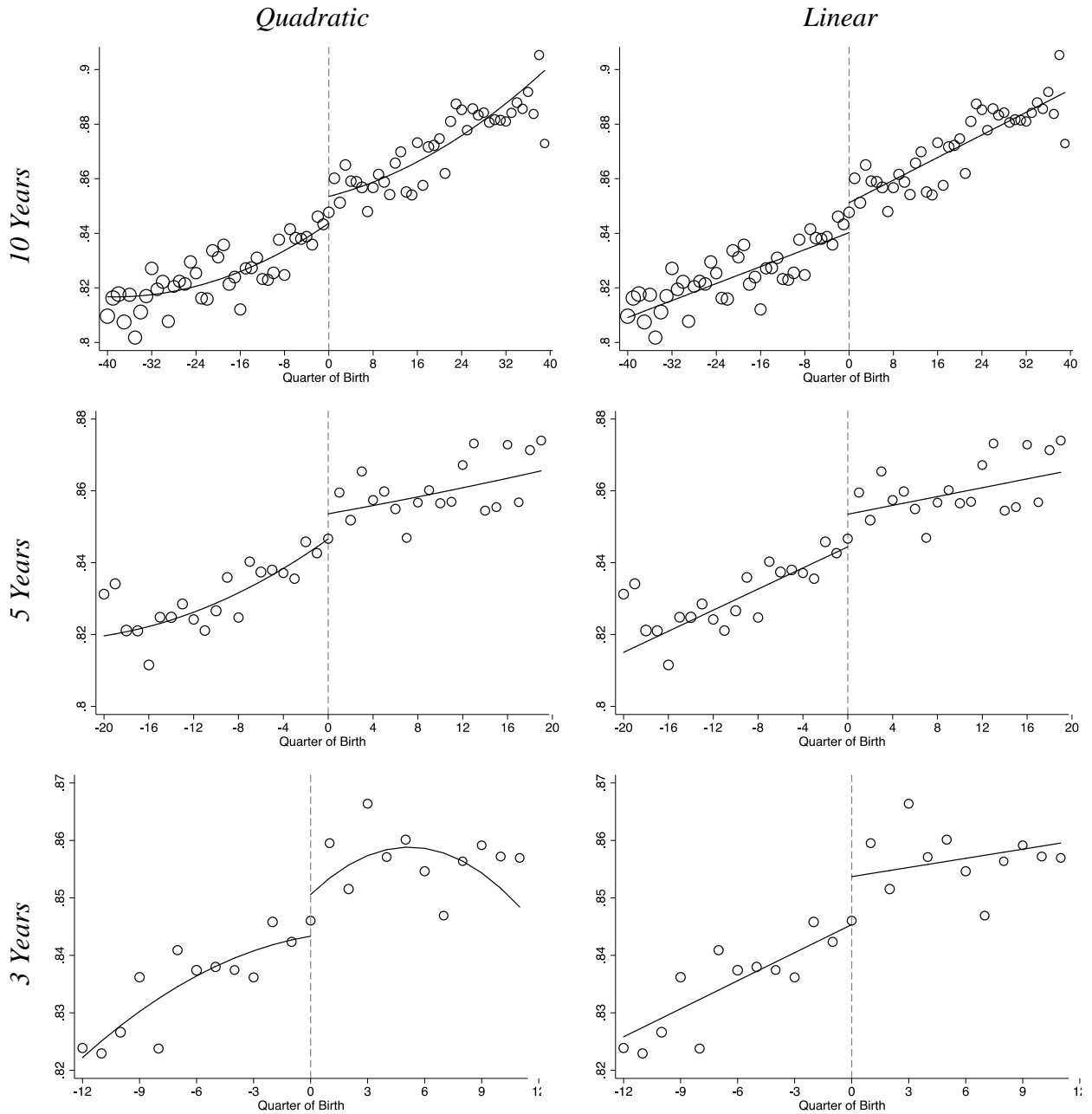# Appendix Figure B9: Distributional Effects on Blood Pressure (No Controls)

Notes: The figure assesses the sensititivity of the distributional effects on the blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the blood pressure index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. No controls.

Appendix Figure B10: Distributional Effects on Blood Pressure (With Controls)

Notes: The figure assesses the sensititivity of the distributional effects on the blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the blood pressure index. The left-hand side column uses quadratic trends in date of birth. The right-hand side column uses linear trends in date of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The regressions include the following set of controls: gender, age in days (at the time of the baseline assessment) and age squared, dummies for ethnicity, dummies for country of birth, and dummies for calendar month of birth.
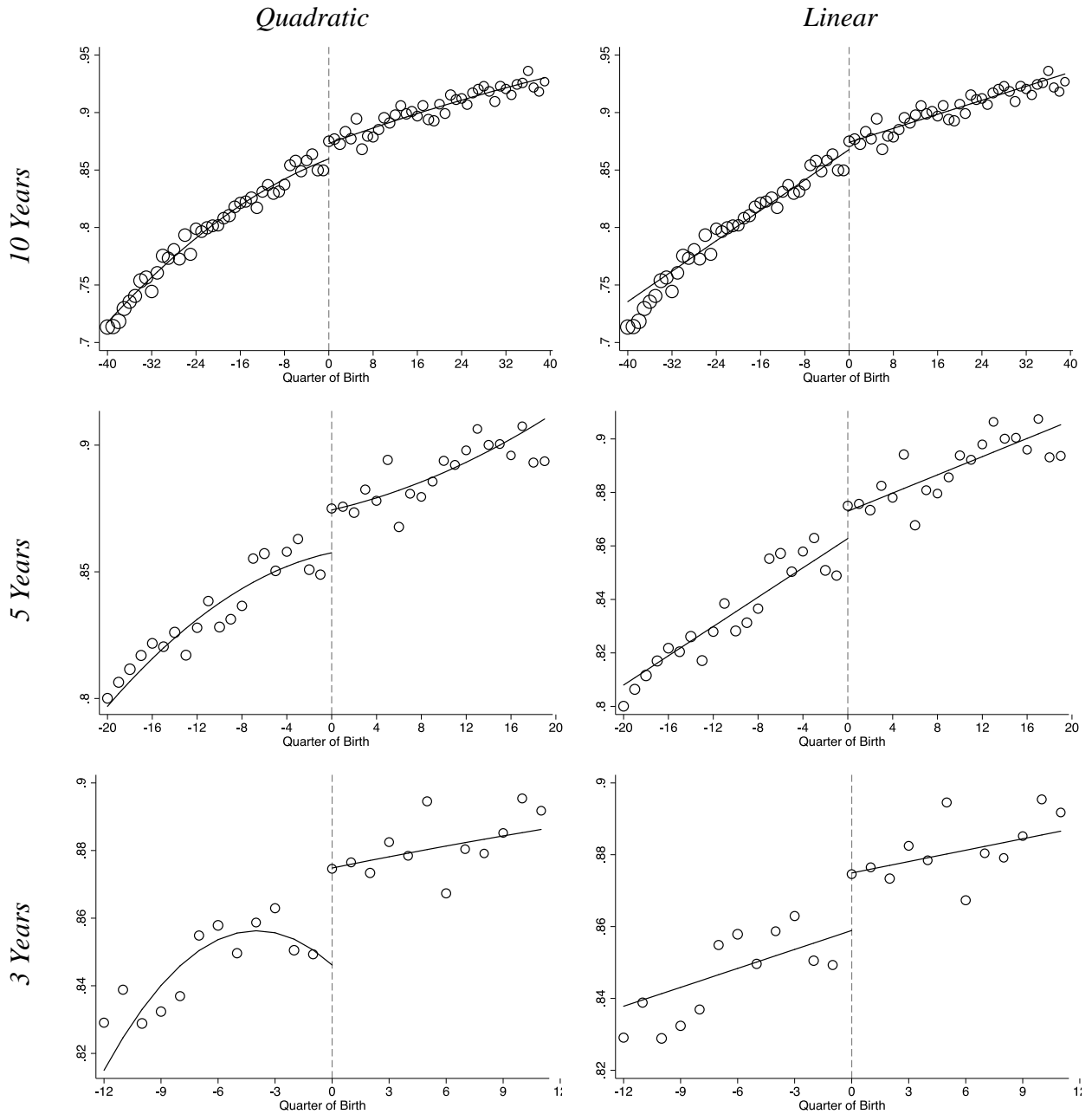
Appendix Figure B11: Fraction with a Body Size Index Smaller than 1 Standard Deviation by Quarter of Birth

Notes: The figure assesses the sensititivity of the results for body size index to the choice of bandwidth and to the use of linear trends. It shows the fraction of study participants with a body size index smaller than 1 standard deviation by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.
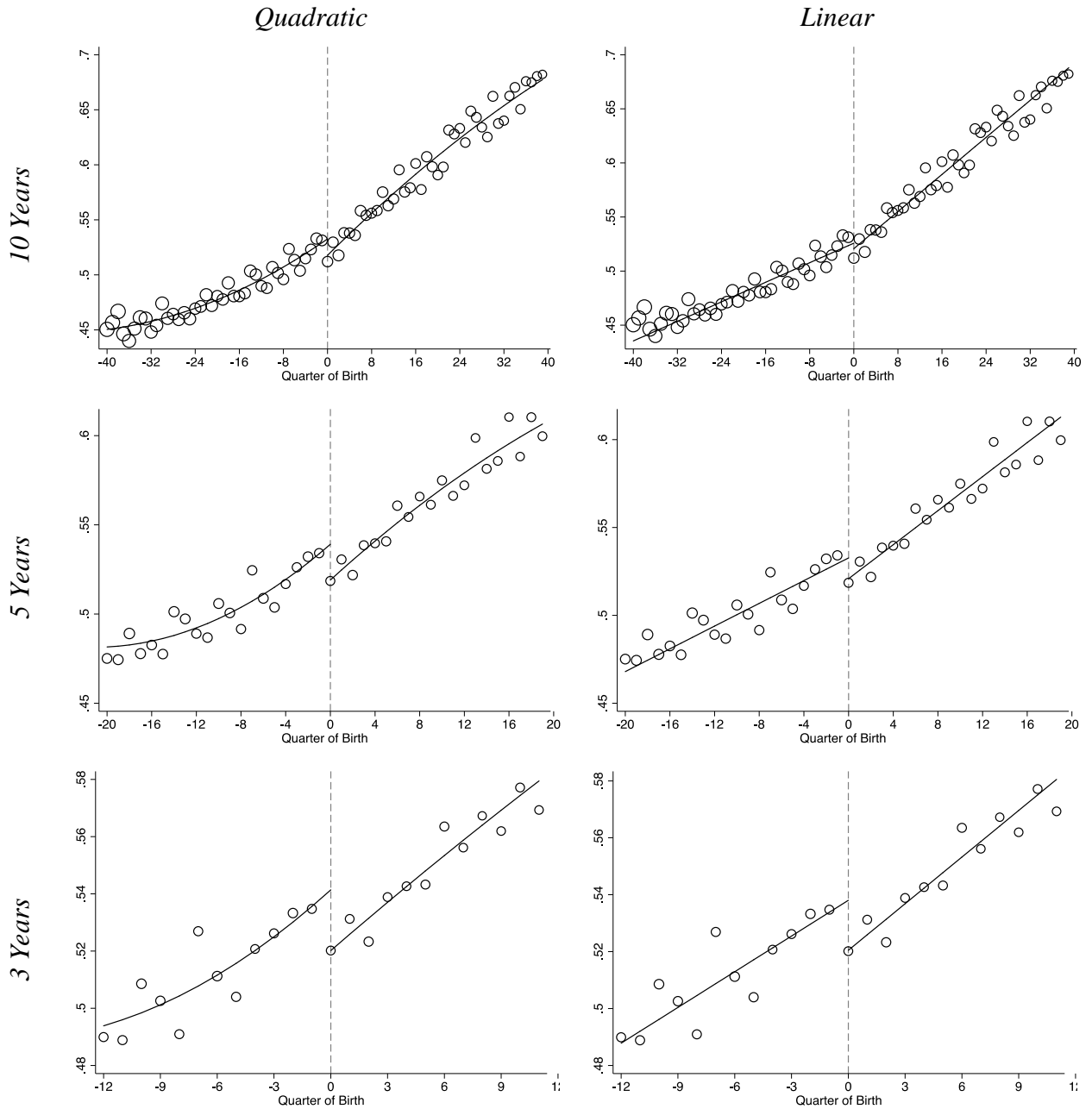
# Appendix Figure B12: Fraction with a Lung Function Index Smaller than 1 Standard Deviation by Quarter of Birth

Notes: The figure assesses the sensitititvity of the results for lung function index to the choice of bandwidth and to the use of linear trends. It shows the fraction of study participants with a lung function index smaller than 1 standard deviation by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

**Appendix Figure B13: Fraction with a Blood Pressure Index Smaller than 0 (Standard Deviations) by Quarter of Birth**

Notes: The figure assesses the sensititivity of the results for blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the fraction of study participants with a blood pressure index smaller than 0 standard deviations by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

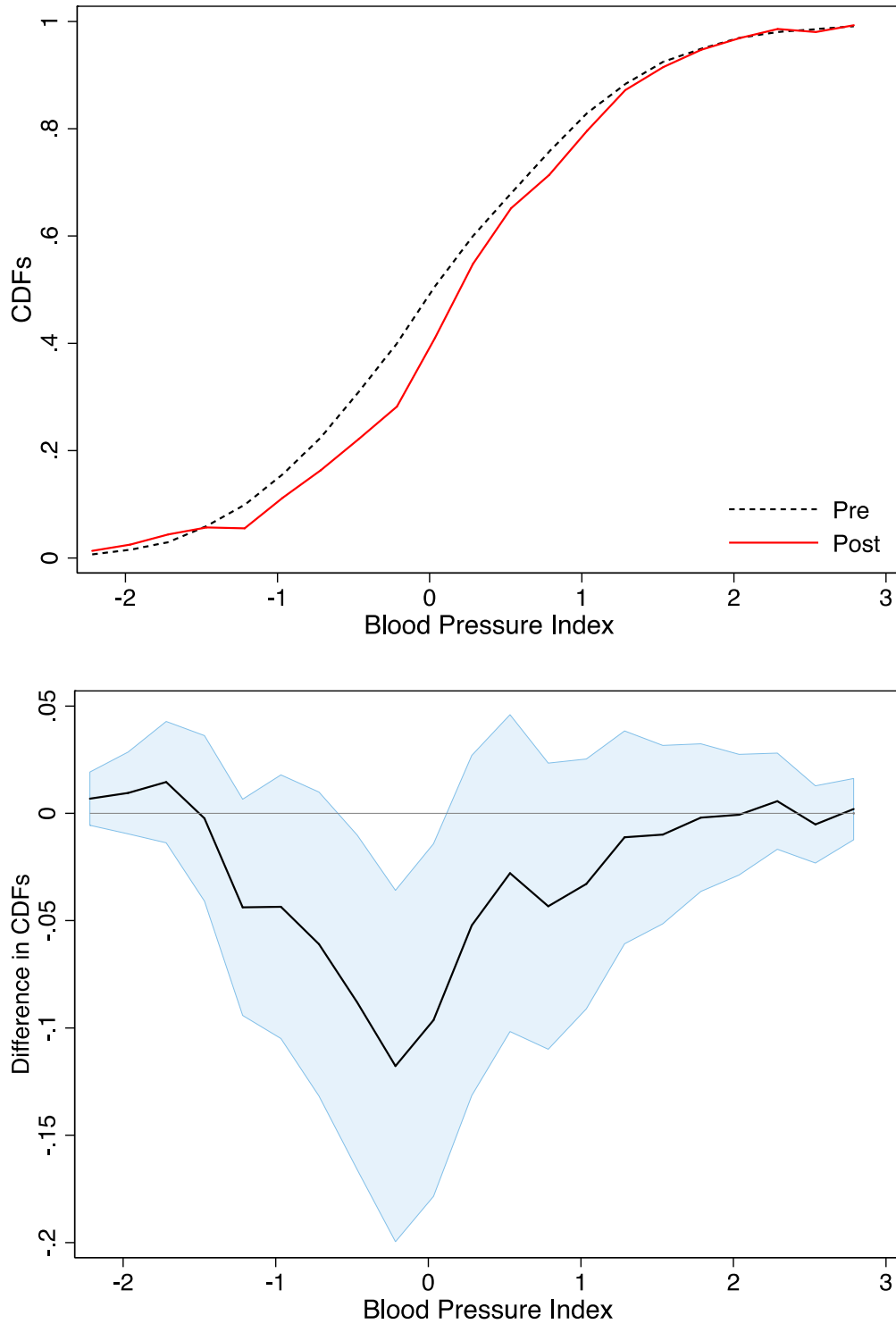**Appendix Figure B14: Distributional Effects on Body Size Index**
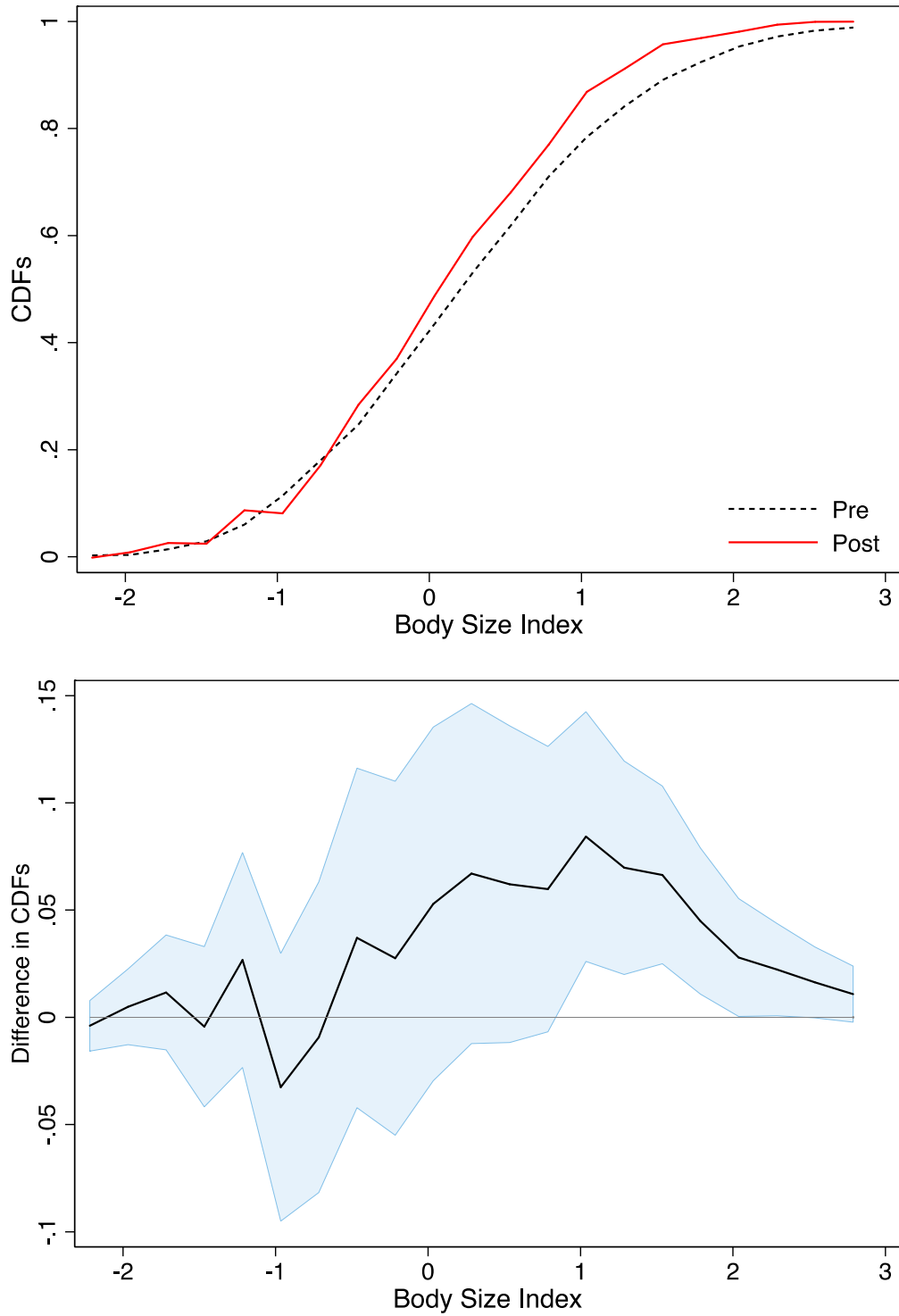
Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 4 in the paper, showing the pre- and post-reform CDF of body size index for compliers. The black solid line in the bottom figure shows the difference between the post- and pre-reform CDFs shown in the top figure. The blue areas show 95% confidence intervals. Inference based on these confidence intervals is problematic because it leads to a large number of highly correlated statistical tests, raising concerns about multiple hypothesis testing.

Appendix Figure B15: Distributional Effects on Lung Function Index
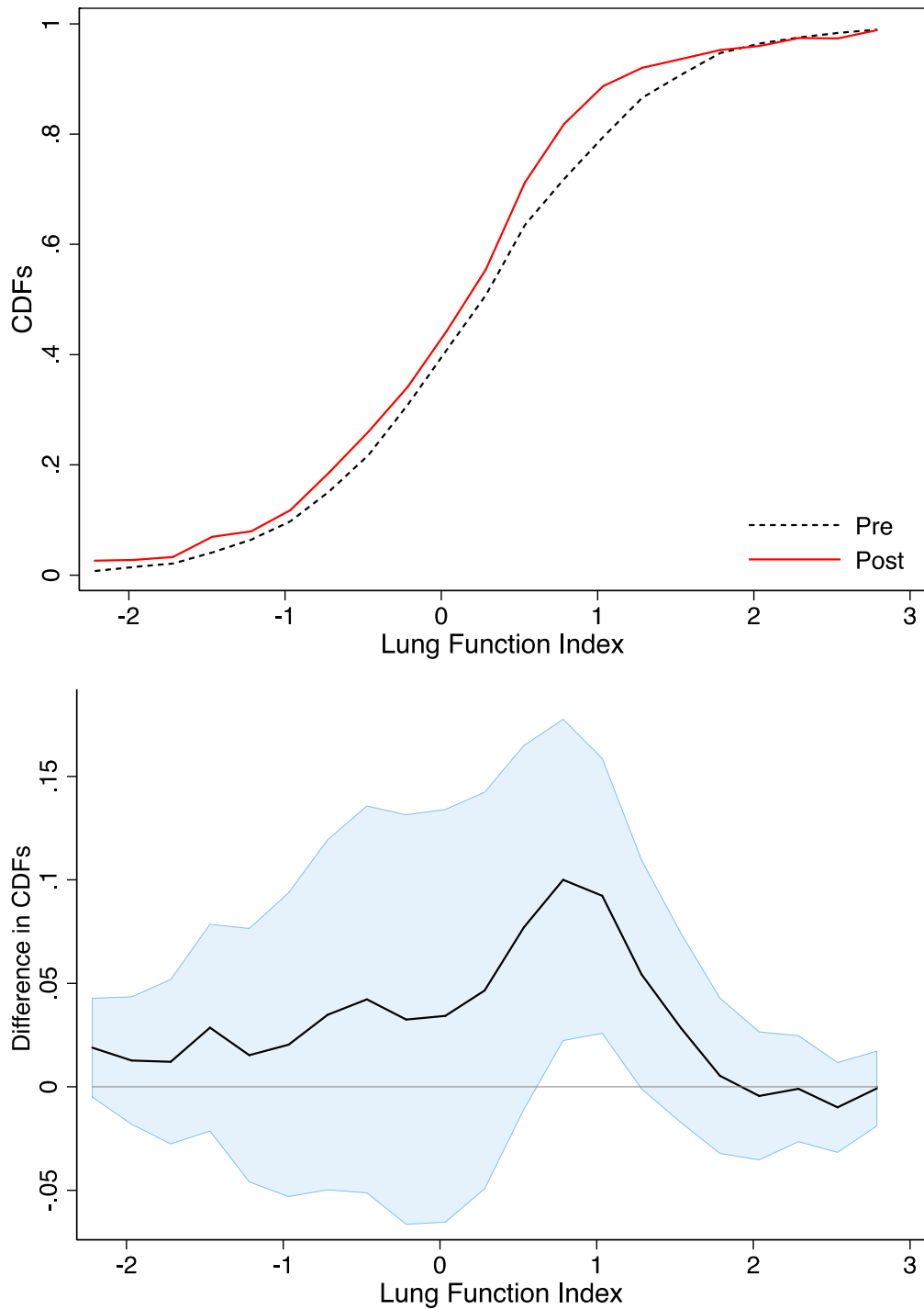
Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 5 in the paper, showing the pre- and post-reform CDF of lung function index for compliers. The black solid line in the bottom figure shows the difference between the post- and pre-reform CDFs shown in the top figure. The blue areas show 95% confidence intervals. Inference based on these confidence intervals is problematic because it leads to a large number of highly correlated statistical tests, raising concerns about multiple hypothesis testing.

Appendix Figure B16: Distributional Effects on Blood Pressure Index

Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 6 in the paper, showing the pre- and post-reform CDF of blood pressure index for compliers. The black solid line in the bottom figure shows the difference between the post- and pre-reform CDFs shown in the top figure. The blue areas show 95% confidence intervals. Inference based on these confidence intervals is problematic because it leads to a large number of highly correlated statistical tests, raising concerns about multiple hypothesis testing.
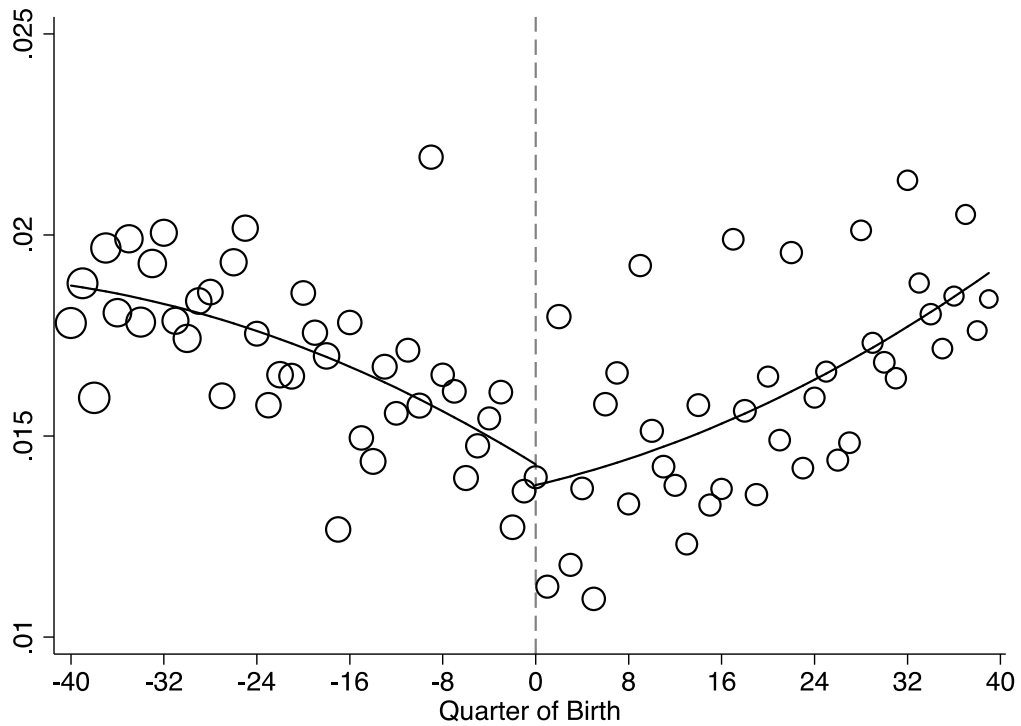
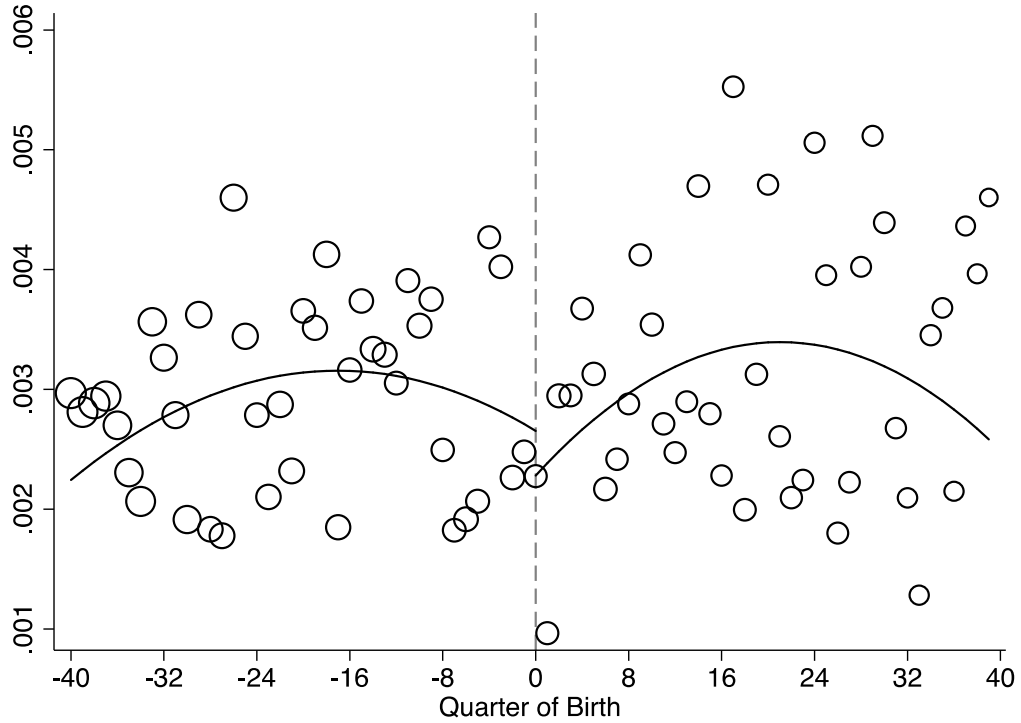# Appendix C

Appendix Figure C1: Missing Body Mass Index
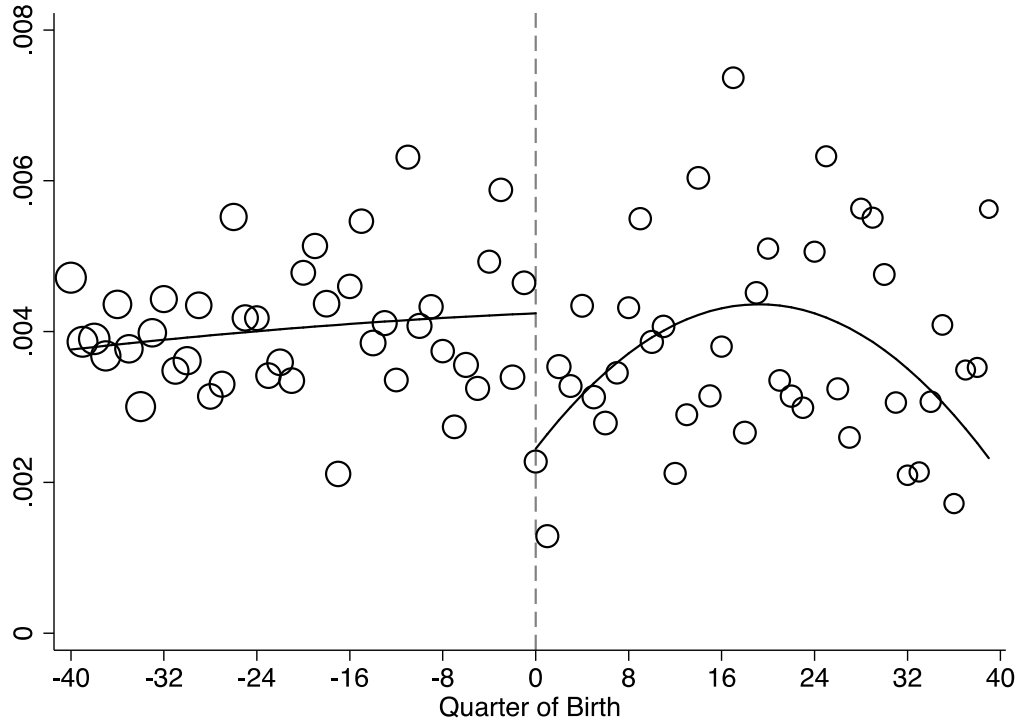
Notes: The figure shows the fraction of study participants for whom data on BMI was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

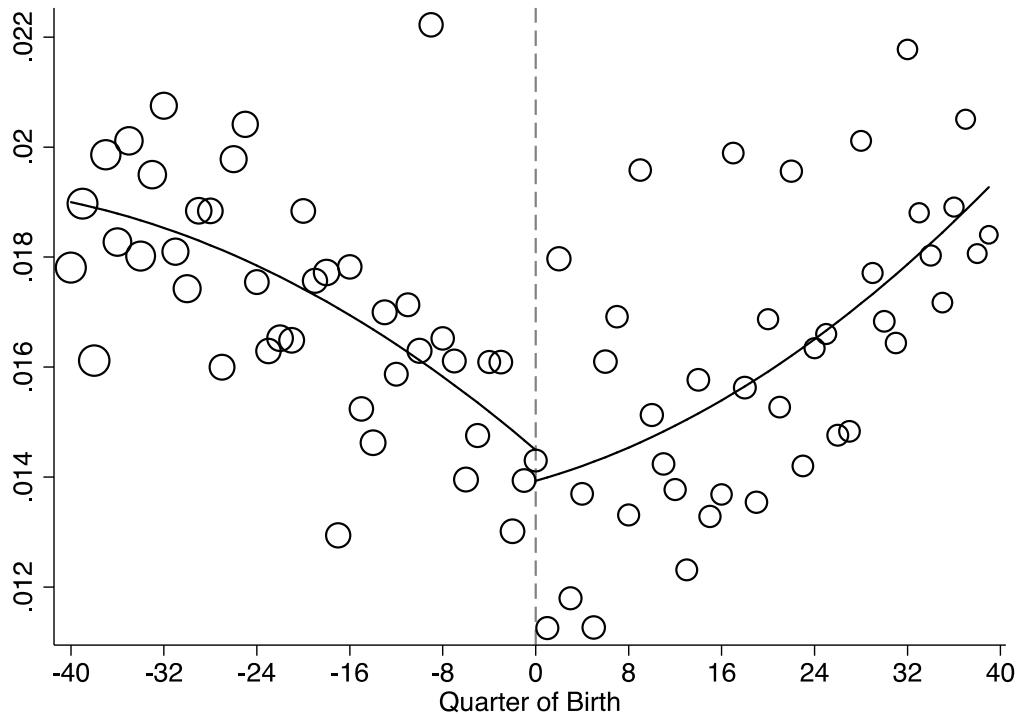## Appendix Figure C2: Missing Body Fat Percentage

Notes: The figure shows the fraction of study participants for whom data on body fat percentage was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure C3: Missing Waist-Hip Ratio

Notes: The figure shows the fraction of study participants for whom data on waist-hip ratio was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure C4: Missing Body Size Index

Notes: The figure shows the fraction of study participants for whom data on body size index was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
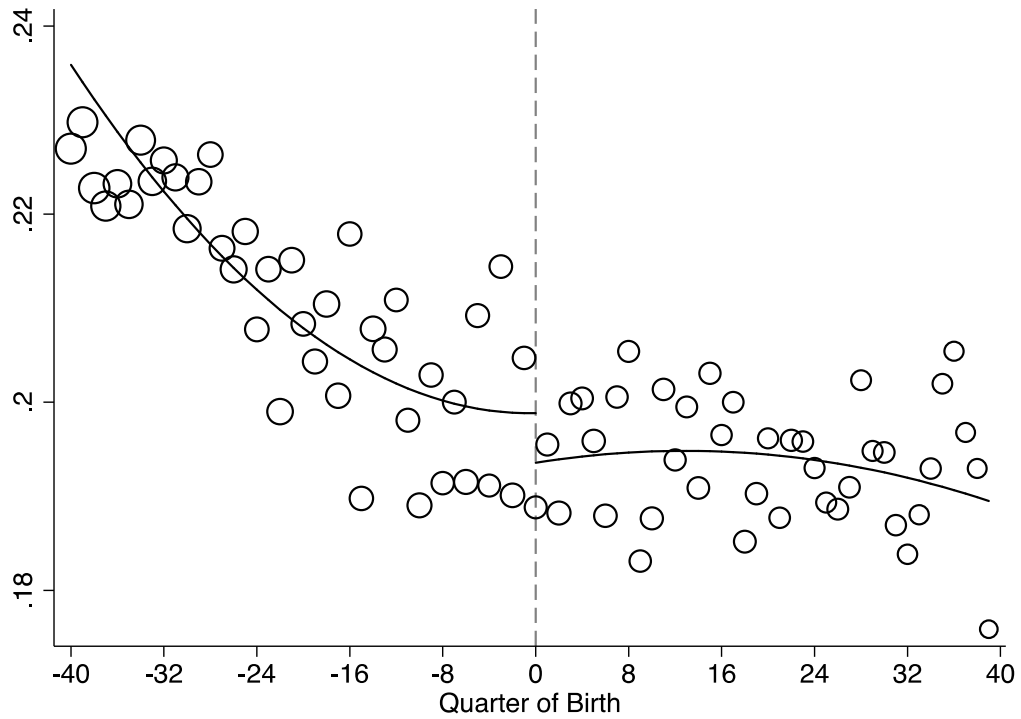
Appendix Figure C5: Missing Lung Function Index

Notes: The figure shows the fraction of study participants for whom spirometry data was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.
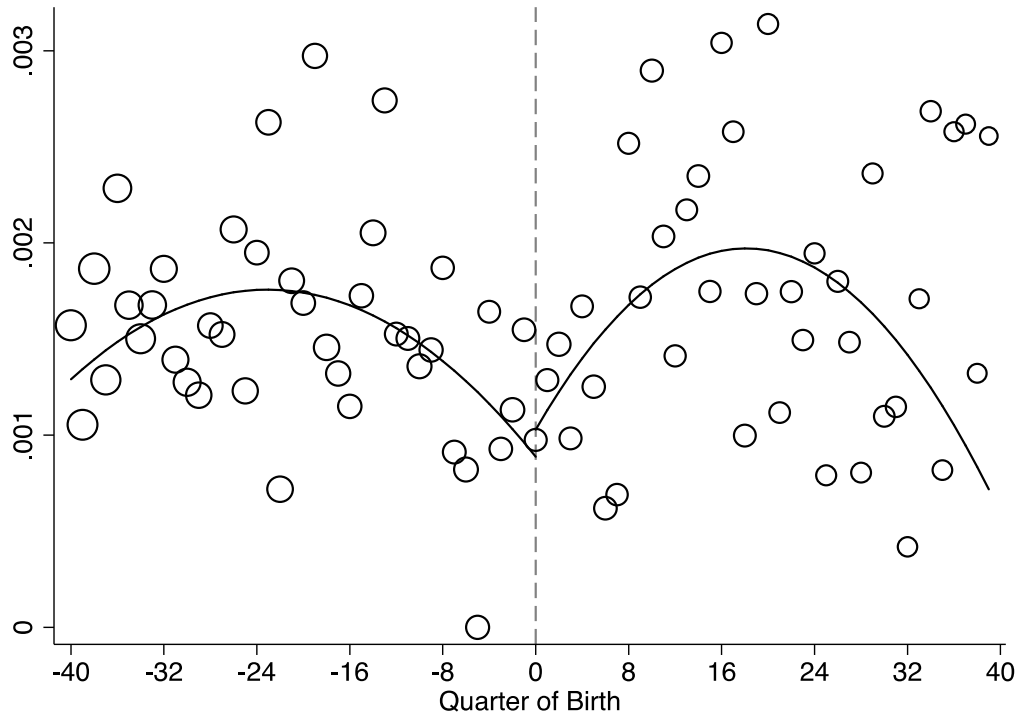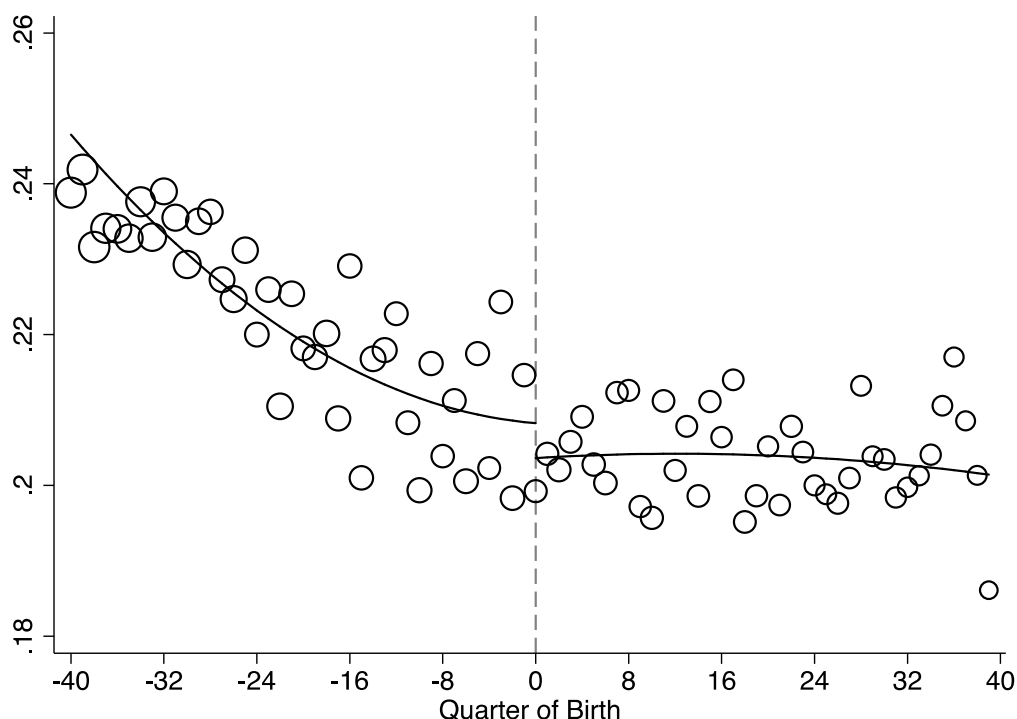
Appendix Figure C6: Missing Blood Pressure Index

Notes: The figure shows the fraction of study participants for whom data on blood pressure was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Figure C7: Missing Summary Index

Notes: The figure shows the fraction of study participants for whom data on the summary index was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic polynomials in quarter of birth that capture birth cohort trends. The circumference of each circle reflects the number of participants born in that quarter. $N = 271,082$.

Appendix Table C1: Missing Outcomes

	BMI		Body Fat Percentage		Waist-hip Ratio		Body Size Index	
Post	-0.002	-0.002	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001
	[0.001]***	[0.001]***	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]
Controls?	N	Y	N	Y	N	Y	N	Y
Mean of Y	0.00467	0.00467	0.0144	0.0144	0.00320	0.00320	0.0147	0.0147

	Lung Function Index		Blood Pressure Index		Summary Index	
Post	-0.006	-0.005	0.000	-0.000	-0.005	-0.005
	[0.005]	[0.005]	[0.000]	[0.000]	[0.005]	[0.005]
Controls?	N	Y	N	Y	N	Y
Mean of Y	0.200	0.200	0.00130	0.00130	0.210	0.210

Notes: The table investigates whether there are discontinuities in missing outcomes at the September 1, 1957 cutoff. It reports the coefficient on an indicator for being born on or after September 1, 1957 (i.e., "Post") from regressions where the dependent variables is listed in the column. The regressions also included quadratic polynomials in date of birth, which were allowed to differ on either side of the cutoff. The mean of Y corresponds to the fraction of study participants born in the 12 months before September 1, 1957 for whom the outcome of interest was missing.
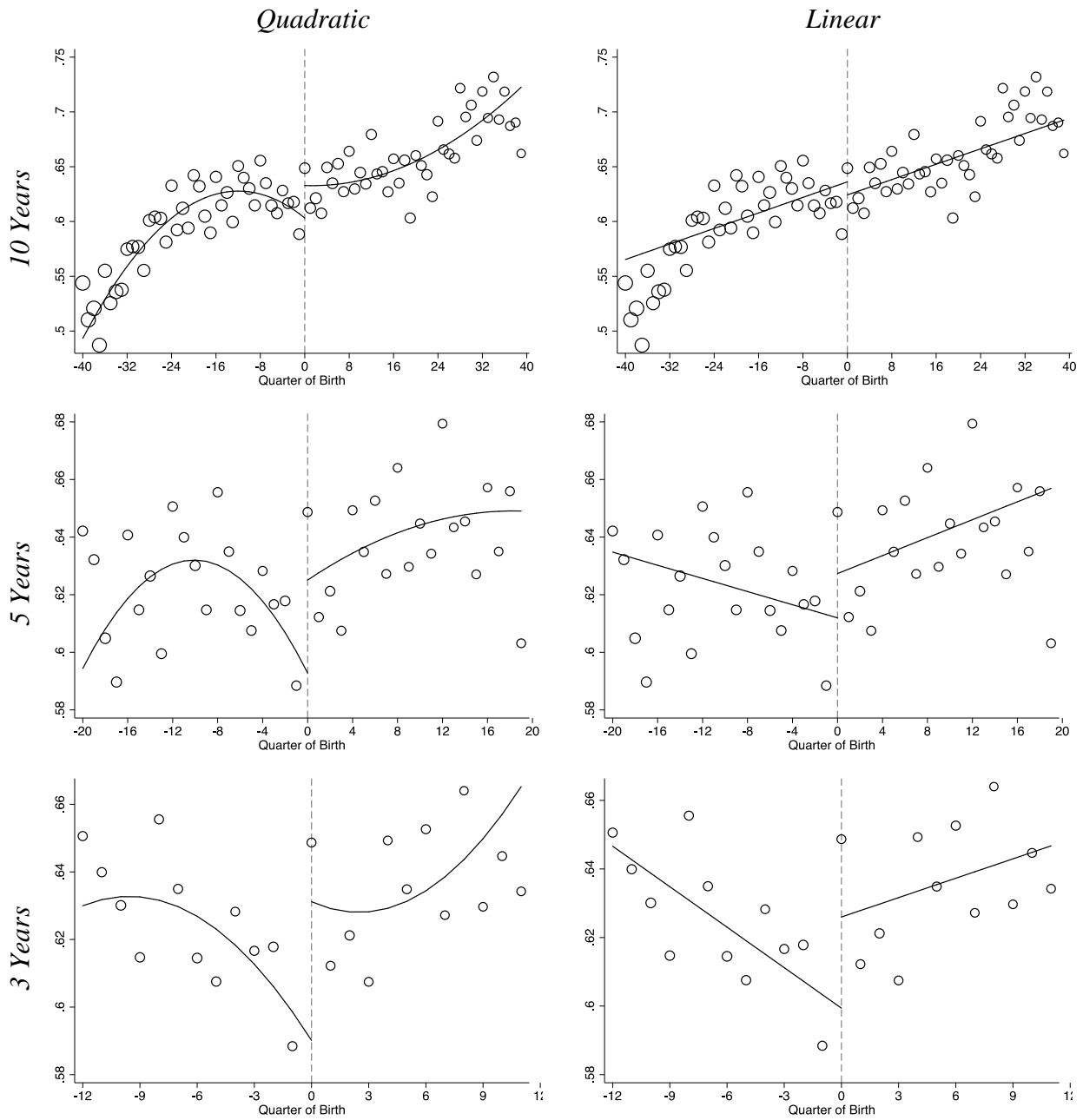
Appendix D

Appendix Figure D1: Map with Locations of 22 Assessment Centers

Notes: The figure shows the location of the 22 assessment centers (as well as the location of the pilot study).

Appendix Figure D2: Fraction Staying in School until Age 17 by Quarter of Birth
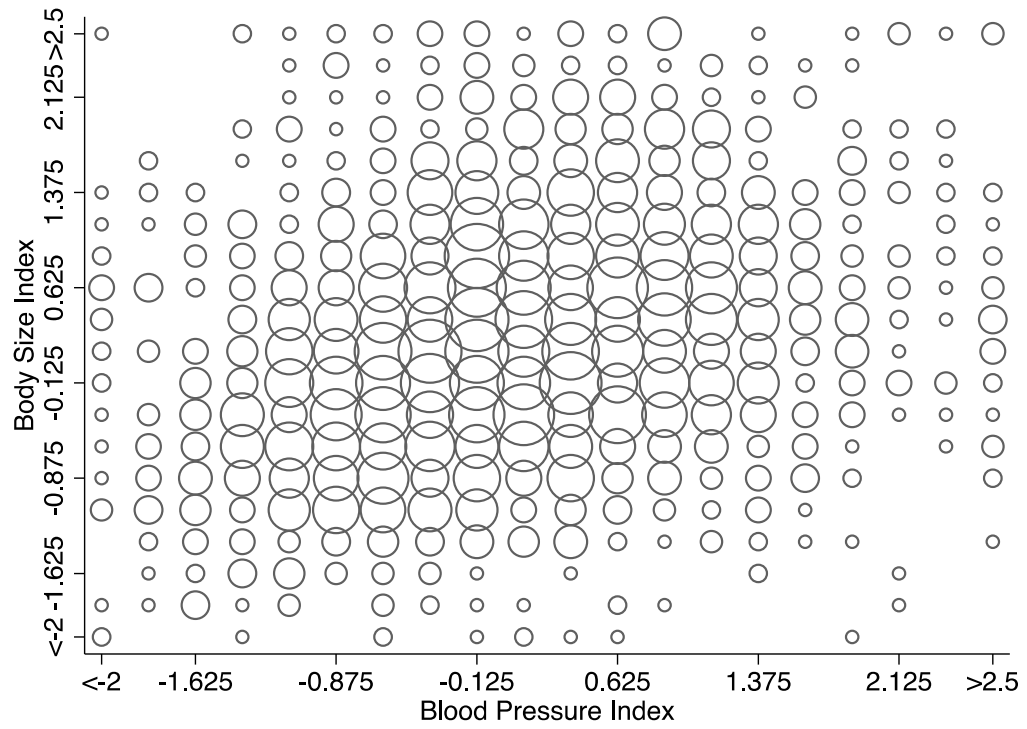

Notes: The figure shows the fraction of study participants who stayed in school until age 17 by quarter of birth for different specifications. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year bandwidth. The middle row uses a 5-year bandwidth. The bottom row uses a 3-year bandwidth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The circumference of each circle reflects the number of participants born in that quarter.

**Appendix Table D1: Effect of 1972 ROLSA
on Fraction Staying in School until Age 17**

| | 10 Years | | | |
	Quadratic		*Linear*	
Post	0.028	0.018	-0.011	-0.017
	[0.006]***	[0.006]***	[0.004]***	[0.004]***
Controls?	N	Y	N	Y
N		271,082		

| | 5 Years | | | |
	Quadratic		*Linear*	
Post	0.027	0.008	0.015	0.005
	[0.009]***	[0.009]	[0.006]**	[0.006]
Controls?	N	Y	N	Y
N		129,222		

| | 3 Years | | | |
	Quadratic		*Linear*	
Post	0.038	0.009	0.024	0.009
	[0.011]***	[0.012]	[0.008]***	[0.008]
Controls?	N	Y	N	Y
N		76,901		

Notes: The table investigates whether the 1972 school-leaving age reform affected the fraction of study participants who stayed in school until age 17. Each cell corresponds to a separate regression of an indicator variable for whether the study participant stayed in school until (at least) age 17 on an indicator variable for whether the study participant was born on or after September 1, 1957 (i.e., "Post"), and quadratic or linear trends in date of birth. The set of controls include gender, age in days (at the time of the baseline assessment) and age squared, dummies for ethnicity, dummies for country of birth, and dummies for calendar month of birth.

Appendix Figure D3: Joint Distribution Function of Body Size and Blood Pressure Indexes

Notes: The figure shows the joint distribution of body size and blood pressure indices among compliers born in the 12 months before September 1, 1957. The circumference of each circle reflects the mass in that interval. $N = 2,210$.

**Appendix Table D2: The categories of the
2000 National Statistics Socio-economic Classification (NS-SEC)**

1	Higher managerial and professional occupations
2	Lower managerial and professional occupations
3	Intermediate occupations
4	Small employers and own account workers
5	Lower supervisory and technical occupations
6	Semi-routine occupations
7	Routine occupations

Notes: The table shows the cateogires of the 2000 National Statistics Socio-economic
Classification (NS-SEC) of occupations.

# Appendix Table D3: Comparison with Clark and Royer (2010) and Davies et al. (2017)

| | Barcellos, Carvalho and Turley | | | | | Clark and Royer (2010) | | | | |
	bpd	high_bpd	BMI	Overweight	Obese	bpd	high_bpd	BMI	Overweight	Obese
Education	1.619	0.012	-0.407	0.003	-0.075	0.985	0.009	0.115	0.027	-0.023
	[0.866]*	[0.035]	[0.420]	[0.039]	[0.033]***	[1.235]	[0.049]	[0.443]	[0.052]	[0.037]
Observations	270,647	270,647	270,019	267,549	267,549	15,097	15,097	18,437	18,437	18,473
mean	82.66	0.23	27.41	0.65	0.25	74.17	0.20	26.28	0.56	0.18

Notes: Instrumental variable estimates of the effect of one extra year of education on health measures. Our results are based on UK Biobank data and use a 10 yrs bandwidth around the September 1, 1957 cutoff. Clark and Royer (2010) use pooled data from the Health Survey of England and a 5 yrs bandwidth.

| | Barcellos, Carvalho and Turley | | | Davis et al. (2017) | | |
	BMI	bps	bpd	BMI	bps	bpd
after	-0.061	0.426	0.243	-0.297	-0.076	-0.127
	[0.063]	[0.213]**	[0.130]*	[0.084]***	[0.466]	[0.361]
Bandwidth	120 mo	120 mo	120 mo	45 mo	50 mo	47 mo

Note: Reduced-form estimates of the effect of the 1972 ROSLA on health. Our estimates are based on a quadratic polynomial on date of birth, 120 months bandwidths and a sample that includes England, Wales and Scotland. Davis et al. (2017) use linear polynomials, CCT optimal bandwidths and include England only. Both papers use UK Biobank data.

Appendix Figure D4: Pre-Reform Cumulative Distribution of Body Size Index for Compliers and for Entire Population
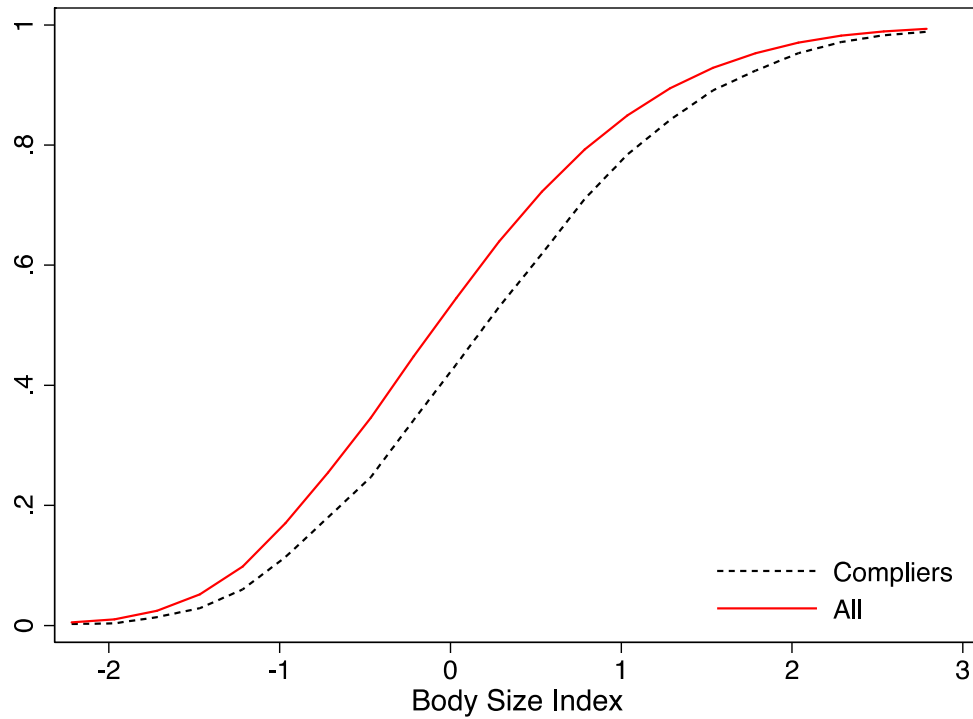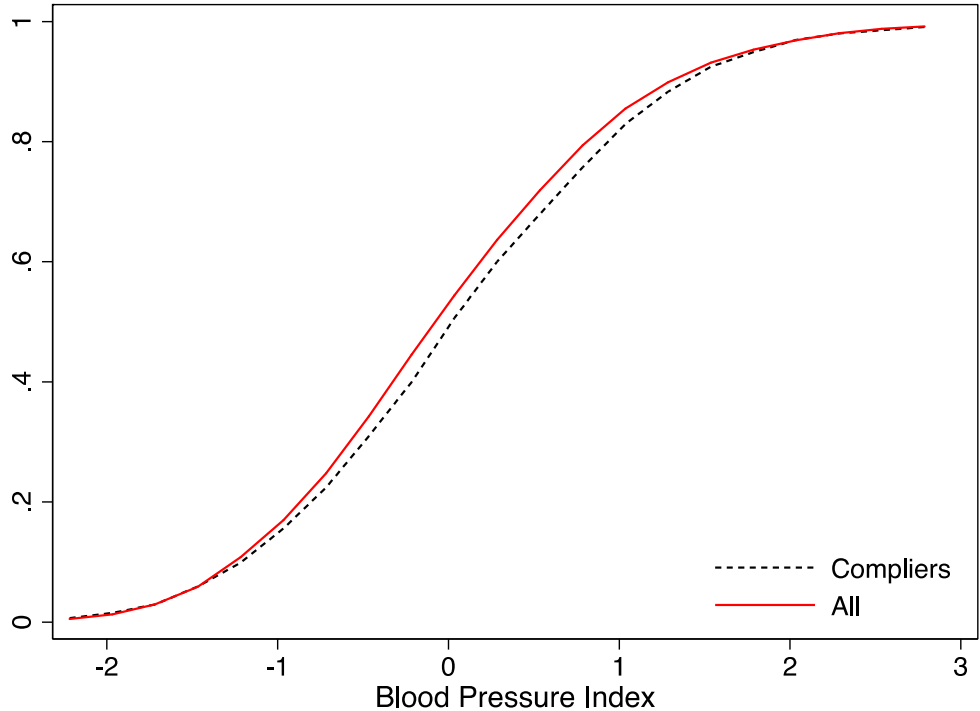

Notes: The figure shows the pre-reform CDFs of body size index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. N = 33,228 (compliers) and 158,707 (all).

Appendix Figure D5: Pre-Reform Cumulative Distribution of Lung Function Index for Compliers and for Entire Population
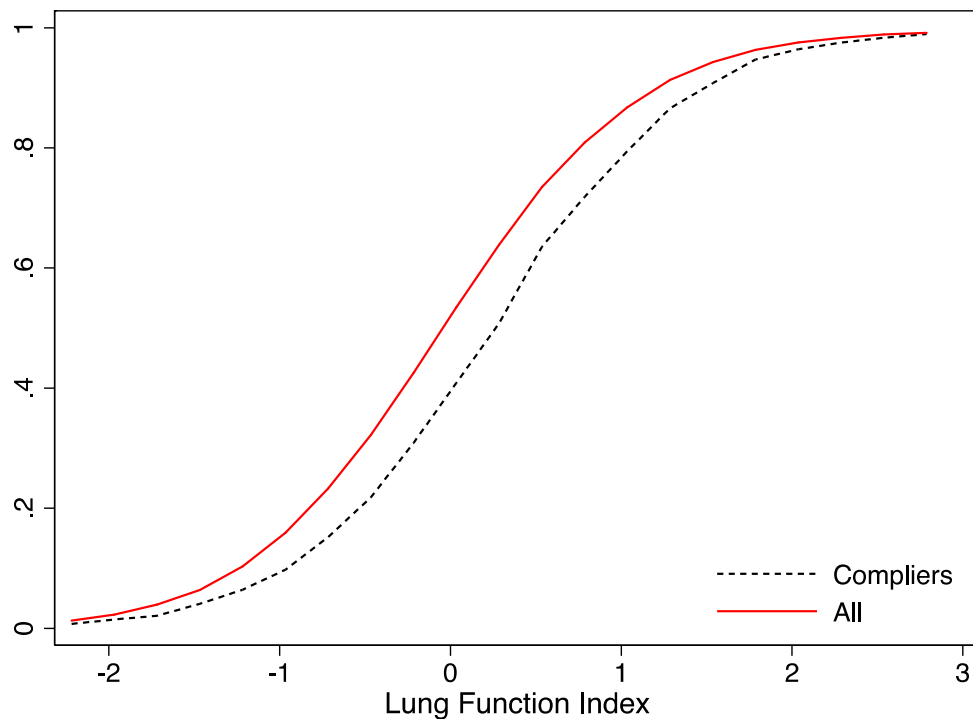
Notes: The figure shows the pre-reform CDFs of lung function index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. N = 25,021 (compliers) and 127,195 (all).

Appendix Figure D6: Pre-Reform Cumulative Distribution of Blood Pressure Index for Compliers and for Entire Population

Notes: The figure shows the pre-reform CDFs of blood pressure index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. $N = 33,882$ (compliers) and 161,264 (all).