NBER WORKING PAPER SERIES

DISTRIBUTIONAL EFFECTS OF EDUCATION ON HEALTH

Silvia H. Barcellos
Leandro S. Carvalho
Patrick Turley

Working Paper 25898
http://www.nber.org/papers/w25898

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019, Revised March 2020

This paper benefited from discussions with James Banks, Dan Benjamin, Damon Clark, Maria Fitzpatrick, Dana Goldman, Mireille Jacobson, Arie Kapteyn, Adriana Lleras-Muney, Neil Mahoney, Michael Mechine, Heather Royer, Jon Skinner, Jim P. Smith, and from the feedback of seminar participants at a number of universities, institutes and conferences. Joao Vilela and Sean Lee provided excellent research assistance. Research reported in this publication was supported through the Roybal Center for Health Decision Making and Financial Independence in Old Age (P30AG024962-13S1 and P30AG024962-13S2), NIA grant K01AG050811-01 (Barcellos), RF1AG055654 (Carvalho), and by the USC Population Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research has been conducted using the UK Biobank Resource under Application Number 15666. It was also conducted using data from The English Longitudinal Study of Ageing. The English Longitudinal Study of Ageing is jointly run by University College London, Institute for Fiscal Studies, University of Manchester and National Centre for Social Research. Genetic analyses have been carried out by UCL Genomics and funded by the Economic and Social Research Council and the National Institute on Aging. All GWAS data has been deposited in the European Genome-phenome Archive. Data governance was provided by the METADAC data access committee, funded by ESRC, Wellcome, and MRC. (2015-2018: Grant Number MR/N01104X/1 2018-2020: Grant Number ES/S008349/1) The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Silvia H. Barcellos, Leandro S. Carvalho, and Patrick Turley. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Distributional Effects of Education on Health
Silvia H. Barcellos, Leandro S. Carvalho, and Patrick Turley
NBER Working Paper No. 25898
May 2019, Revised March 2020
JEL No. I10,I20

ABSTRACT

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 9/1/1957. Using a regression discontinuity design and objective health measures for 0.27 million individuals, we find that education reduced body size 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 stage 2 hypertension.

Silvia H. Barcellos
Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA 90089
and NBER
sbarcell@usc.edu

Leandro S. Carvalho
Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA 90089
leandro.carvalho@usc.edu

Patrick Turley
Massachusetts General Hospital
Richard B. Simches Research Center
185 Cambridge Street
CPZN-6818
Boston, MA 02114
paturley@broadinstitute.org

A data appendix is available at http://www.nber.org/data-appendix/w25898

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 9/1/1957. Using a regression discontinuity design and objective health measures for 0.27 million individuals, we find that education reduced body size 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 stage 2 hypertension. JEL codes: I10, I20.

There is broad disagreement about whether education has a causal effect on health (Grossman 2015). Some studies find that more education causes better health (e.g. Lleras-Muney 2005) while others find little to no effect (e.g. Clark and Royer 2013). Galama, Lleras-Muney, and Kippersluis (2018) argue that heterogeneity in these effects may underlie these conflicting results. Their model lays out conditions for such heterogeneity to exist, and evidence indicates that such conditions are met: (i) heterogeneity in the labor market returns to education (Card 2001); (ii) heterogeneity in the non-market returns to education (Barcellos,

* Barcellos: University of Southern California, Center for Economic and Social Research, 635 Downey Way, Los Angeles, CA 90089-3332 (silvia.barcellos@usc.edu); Carvalho: University of Southern California, Center for Economic and Social Research, 635 Downey Way, Los Angeles, CA 90089-3332 (leandro.carvalho@usc.edu); Turley: Massachusetts General Hospital, Richard B. Simches Research Center, 185 Cambridge Street, CPZN-6818, Boston MA 02114 (paturley@broadinstitute.org). This paper benefited from discussions with James Banks, Dan Benjamin, Damon Clark, Maria Fitzpatrick, Dana Goldman, Mireille Jacobson, Arie Kapteyn, Adriana Lleras-Muney, Neil Mahoney, Michael Mechine, Heather Royer, Jon Skinner, Jim P. Smith, and from the feedback of seminar participants at a number of universities, institutes and conferences. Joao Vilela and Sean Lee provided excellent research assistance. Research reported in this publication was supported through the Roybal Center for Health Decision Making and Financial Independence in Old Age (P30AG024962-13S1 and P30AG024962-13S2), NIA grant K01AG050811-01 (Barcellos), RF1AG055654 (Carvalho), and by the USC Population Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research has been conducted using the UK Biobank Resource under Application Number 15666. It was also conducted using data from The English Longitudinal Study of Ageing. The English Longitudinal Study of Ageing is jointly run by University College London, Institute for Fiscal Studies, University of Manchester and National Centre for Social Research. Genetic analyses have been carried out by UCL Genomics and funded by the Economic and Social Research Council and the National Institute on Aging. All GWAS data has been deposited in the European Genome-phenome Archive. Data governance was provided by the METADAC data access committee, funded by ESRC, Wellcome, and MRC. (2015-2018: Grant Number MR/N01104X/1 2018-2020: Grant Number ES/S008349/1) The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Carvalho, and Turley 2018); and (iii) heterogeneity in skill formation as a result of education (Grenet 2013).

This paper investigates whether the effects of education on health are heterogeneous by studying how education affects the distribution of health. If education affects some parts of the health distribution but not others, then the effect of education on health must vary across individuals. In particular, the effect may be larger for individuals with poorer health. In a Grossman-type model, education will cause the unhealthy to invest more in their health than the healthy because of the diminishing marginal utility of health (Kippersluis and Galama 2014).[1]

Since little is known *ex-ante* about the dimensions along which such effect may vary, a distributional approach allows for an unsupervised investigation of heterogeneity, avoiding multiple hypotheses testing and concerns about the lack of good measures of relevant dimensions. Moreover, if the effects are concentrated at particular parts of the distribution, a distributional test may be better powered than a test of difference in means—see Appendix F.

For this purpose, we exploit a well-known schooling reform in the United Kingdom using a regression discontinuity design. 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." The reform had the primary effect of inducing students who would have dropped out at age 15 to obtain exactly one additional year of education. We estimate the distributional effects of education on health from discontinuities in the cumulative distribution function (CDF) of health at the birthdate cutoff.

[1] Consistent with this hypothesis, in related work we show that the additional schooling had larger effects on the BMI of those with greater genetic risk of obesity (Barcellos et al. 2018).

We use data from the UK Biobank, a study that collected multiple *objective* and *continuous* measures of health between 2006 and 2010; 34-38 years after the policy change. Using standardized protocols, healthcare technicians and nurses measured the BMI, body fat percentage, waist and hip circumferences, lung function, and blood pressure of more than a quarter million people born in England, Scotland, and Wales between September 1, 1947 and August 31, 1967. The unprecedented availability of objective health measures for such a large sample permits well-powered estimation of distributional treatment effects. To ease concerns about multiple hypotheses testing, we focus our analysis on three health indices constructed from the multiple measures available: body size, lung function, and blood pressure.

There are three main takeaways from the distributional analysis. First, staying in school until age 16 improves one dimension of health—there is a reduction in body size—and *worsens* another: there is an increase in blood pressure.[2] Second, it reveals that these effects vary considerably along the health 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 effect on average body size. Third, the effects on body size and on blood pressure occur in different parts of their respective distributions: while the effect on body size is 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 bottom and top halves of the CDFs of our three health indices. The test rejects at the 5% significance level the

[2] In our sample, there are very few participants with low blood pressure or who are underweight.
[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. See Appendix Figure D3.

null of equality for the top half of the body size distribution (p-value of 0.013) and for the bottom half of the blood pressure distribution (p-value of 0.010). We only find suggestive evidence of a positive effect on lung function.

The results from our distributional analyses suggest that the effects of education on health are heterogeneous. Our estimates recover the distributional effects on compliers, that is, students who stayed in school an additional year because of the schooling reform. If the effects were homogeneous, we would see effects across the whole support of the complier distribution, with the largest effect at the mode (see Appendix I). This is because, if education affects each complier equally, the distributional effect at a given point should roughly correspond to the mass at this point in the pre-reform complier distribution. As described above, this is not what we find: effects are concentrated at the top of the body size distribution, with the largest effect at the 78th percentile. Therefore, our findings are not consistent with homogeneous treatment effects. Importantly, our method recovers differences in the health distribution of compliers only and does not rely on rank preservation assumptions.

We also conduct an exploratory analysis to investigate possible channels through which education may affect health. While improvements in SES and diet may explain the reduction in body size, the channels for the increase in blood pressure are less clear. We do show, however, that education allowed individuals to hold better jobs in middle age. One possibility is that, by changing the types of occupations and careers individuals have, education might have had 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) and with findings that, despite improving SES outcomes, the 1972 reform had detrimental effects on mental health (Aventano et al. 2017).

For the sake of comparison, we also estimate the effects of education on average health. A number of studies have exploited changes in compulsory schooling laws to study such effects (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; Janke et al. 2018; Meghir et al. forthcoming). We find that staying in school until age 16 reduces body size by 0.15 of a standard deviation and increases blood pressure by 0.15 of a standard deviation, but these estimates are only significant at the 10% level.[4]

In an apparent contrast to the results of this paper, Clark and Royer (2013) estimate a precise null effect of the 1972 ROSLA reform on mortality. They also find no statistically significant effect on the averages of non-mortality health outcomes.[5] Our estimates of the effects on average health are, however, broadly consistent with their corresponding estimates. While Clark and Royer had very large samples to estimate the effect of education on mortality, their estimates of the effects on non-mortality health outcomes are less precise because they had more limited samples for such outcomes. For this reason, our estimates lie within the 95% confidence intervals of Clark and Royer (2010, 2013)'s corresponding estimates.

While the main take-away from Clark and Royer (2013) is that education does not affect health, our results suggest a more nuanced story. The distributional analysis shows that education can affect the health of some subpopulations. The heterogeneity of the effects of education on health may explain the conflicting findings in the literature (Galama et al. 2018). First, education affects different parts of the health distribution differently, implying that the effects vary across individuals. Second, the effects also vary across outcomes: the additional schooling caused a reduction in body size and an increase in blood pressure.

[4] We also find an improvement in lung function that is marginally significant at 10% but that goes away once controls are added.

[5] Clark and Royer also analyzed the effects of education on particular points of the distribution of BMI and blood pressure, but these clinical cutoffs are arguably arbitrary (Komaroff 2016) and the approach may overlook changes in other parts of the distribution.

Despite this, the question remains as to how education could affect some health outcomes while having no downstream effects on mortality. There are several hypotheses for such a result. First, those affected by the 1972 ROSLA would have been less than 50 years old at the time of the Clark and Royer study. It is possible that not enough time had passed to observe meaningful differences in mortality.[6] Second, health is a multi-dimensional construct, and, as we see in this study, education may improve some dimensions of health while worsening others. As a result, the ultimate effect of education on mortality may be attenuated by these opposing factors. In short, Clark and Royer's finding of no effect on mortality are plausibly consistent with our results.

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.

[6] This argument would not apply to their 1947 ROSLA results, but we also do not know whether the 1947 ROSLA had the same effects on health as we estimate for the 1972 ROSLA.

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 it to 16 years when conditions allowed. The Minister did so in January 1972 for Scotland (Statutory Instrument No. 59)[7] and in March 1972 for England and Wales (Statutory Instrument No. 444)[8]. 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. 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. We estimate that the policy increased the fraction of UK Biobank participants who stayed in school until age 16 by 15 percentage points – see Table 1.[9]

[7] http://www.legislation.gov.uk/uksi/1972/59/pdfs/uksi_19720059_en.pdf
[8] http://www.legislation.gov.uk/uksi/1972/444/pdfs/uksi_19720444_en.pdf
[9] The effect on staying in school until at least age 17 is an order of magnitude smaller: about 2 percentage points. In Clark and Royer's words (AER 2013): "the 1972 change had small, at best, effects on the fractions completing 11 or fewer years…To a first approximation therefore, one can view these law changes as forcing students that would previously have left at the earliest opportunity to stay in school for one more year" (pg. 2102).

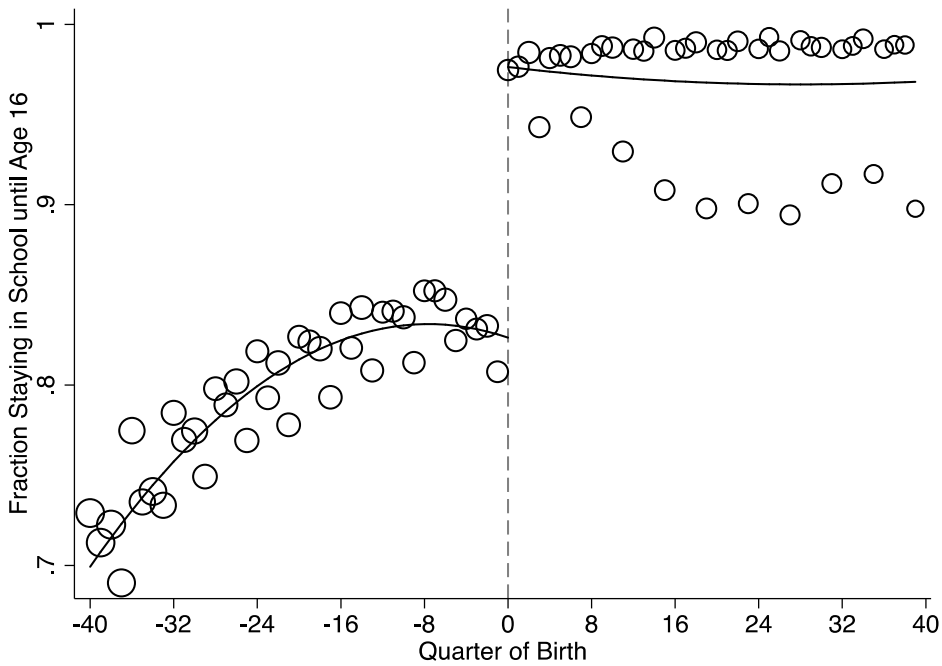


Figure 1: Fraction Staying in School until Age 16 by Quarter of Birth

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.

Notice there is a cyclical drop in the fraction of students staying in school until age 16, corresponding to those born between June and August. This phenomenon is not specific to our data and has been noted by others. According to Clark and Royer (2013), "when the minimum leaving age became age 16, students had to stay until part way through grade 10. Grade 10 finishes with the "O level" exam period and, technically, students finish when they complete their last exam. Since the exam period starts in late May and finishes in mid-June, starting in 1972, students born in late June, July, and August could leave at 15, technically younger than the minimum leaving age (16)" (pg. 2 of Online Appendix). We include calendar month of birth dummies in our regressions to control for this seasonality.

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.[10] 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 A.[11]

Physical measures, such as anthropometrics, spirometry, and blood pressure, were collected of survey participants. The collection was standardized across centers and was conducted by trained nurses or healthcare practitioners. 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*.[12] Weight and body size can be affected by education through diet and physical exercise. Lung

[10] The NHS has contact details for an estimated 98% of the UK population.
[11] When the predetermined variable is continuous, we test whether the pre- and post-reforms distributions are different. When the predetermined variable is dichotomous, we conduct a test of difference in means.
[12] The other objective and continuous measures of health currently available in the UK Biobank are either available for just a subsample (e.g., arterial stiffness, bone densitometry of heel, ECG) or there is no clear hypothesis on how they could be affected by education (e.g., hand grip strength).

function can be affected by education through smoking and pollution. Blood pressure may be affected by education through diet, physical exercise, medication, and stress. Moreover, these health dimensions are risk factors for high-prevalence diseases such as diabetes (Chan et al. 1994), lung cancer (Tockman et al. 1987) and cardiovascular disease (Kannel 1996). 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: 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.

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 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.

[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. We can get very similar results if we use exclusively the first measure of BMI (i.e., the weight in kilograms divided by height in meters squared). 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.

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]

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.

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), weighting 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 function and the blood pressure indices is 0.10. The correlations between the summary index

[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.

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:

$$Educ_i = a_0 + a_1 Post_i + f(DoB_i) + \boldsymbol{x}_i' \boldsymbol{a}_2 + \varepsilon_i, \tag{1}$$

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 $\boldsymbol{x}_i$ 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 restrict the data to study participants born in England, Scotland, or Wales 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)).[17] In Appendix B we show our main results are robust to the choice of bandwidth and to the use of linear trends.[18] We use triangular kernel weights that give greater weight to

[16] The inclusion of predetermined controls in equation (1) is not needed for identification but can improve the precision of estimates.
[17] We opt to use 10-year bandwidths because, to our knowledge, there are no existing methods to estimate the optimal bandwidth for distributional RDD.
[18] Gelman and Imbens (2016) caution against the use of higher order polynomials (higher than 2) in RDD.

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).[19]

Notice that even though previous work studying the 1972 ROSLA clustered standard errors by month-year of birth (e.g., Clark and Royer 2013; Davies et al. 2017), we do not need to cluster our standard errors because our data include exact date of birth. As Card and Lee (2008) discuss, in applications where the running variable is only reported in coarse intervals (e.g., month-year of birth), researchers have to choose a particular functional form for the model relating the outcomes of interest to the running variable. The deviation between the expected value of the outcome and the predicted value from a given functional form is modeled as a random specification error, which is incorporated in inference by clustering the standard errors for different values of the running variable. This specification error should be negligible in our context because our data include day-month-year of birth. Appendix Table D4 shows that we get virtually identical standard errors estimates irrespective of whether we cluster by date of birth or not.

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.

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

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.[20] Studies using nationally representative data, such as Clark and Royer (2013), estimate this figure to be closer to 25 percentage points. This difference is likely due to the composition of the UK Biobank sample, which is more educated than the overall population (despite the selectivity of the UK Biobank sample, our estimates have internal validity because there is no differential selection on the two sides of the September 1, 1957 cutoff – see Appendix A). One consequence is that the standard errors of our two stages least squares (2SLS) estimates will be *ceteris paribus* larger than of studies with nationally representative data, something that is compensated by the larger sample size of the UK Biobank.

Table 1: Effects on Education

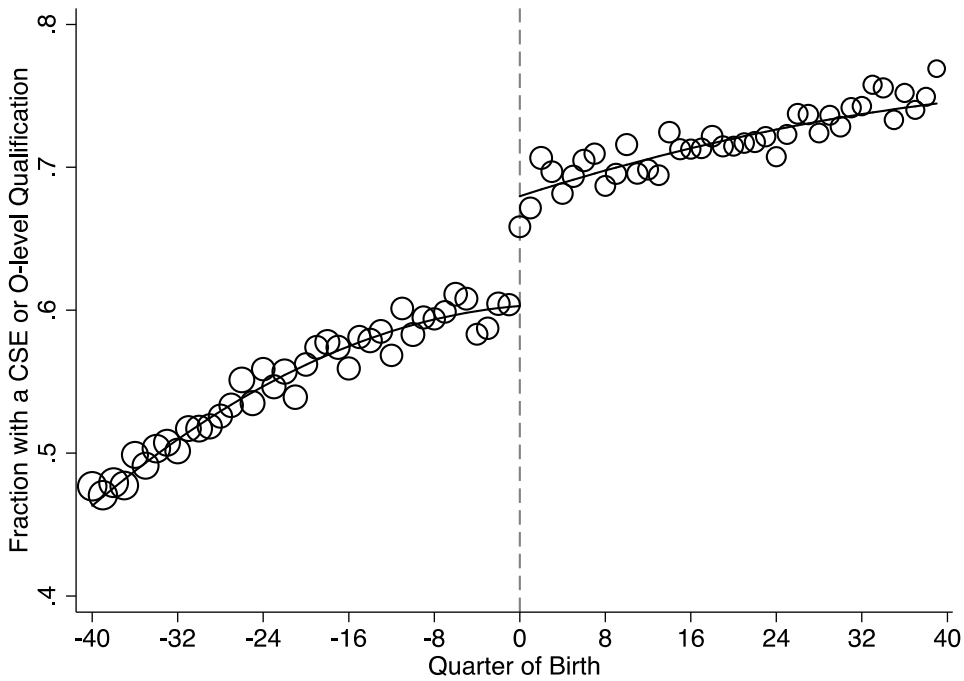
	Left school at age ≥ 16		*No qualification*		*CSE*	
Post	0.150	0.139	-0.048	-0.050	0.059	0.070
	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.005]***	[0.005]***
Controls?	No	Yes	No	Yes	No	Yes
Mean of Y	0.827		0.113		0.205	
	O-level		*A-level*		*College degree*	
Post	0.038	0.035	0.016	0.015	-0.003	-0.005
	[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.

[20] 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 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 reduced-form estimates are statistically significant at 1%. We find no effect of the policy on having a college degree. Consistent with these results, we document that the policy increased household income[21], especially at lower income levels, and enabled workers to get "better jobs", that is, to have occupations with higher socioeconomic status[22]– see Appendix Table E1 and Appendix Table E2.

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:

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

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:

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

[21] These results are broadly consistent with Grenet (2013) that finds that the extra year of schooling induced by the 1972 ROSLA increased wages.

[22] Respondents who were employed or self-employed were asked in a verbal interview to describe their jobs. Respondent's answers were coded following the Standard Occupation Classification 2000. We classify the socioeconomic status of the occupations using the 2000 National Statistics Socio-economic Classification (NS-SEC), the primary social classification in the United Kingdom. See Appendix Table D2.

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. Since the UK Biobank genotyped the full sample, we also test for the smoothness of a pair of genetic variables, which are determined at conception and are objectively measured.[23] Our analyses indicate that these characteristics are smooth across the September 1 1957 threshold, 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: 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).

[23] See Online Appendix of Barcellos, Carvalho and Turley (2018) for details about the construction of genetic variables.

Table 2 shows regression estimates of the effects of the 1972 compulsory schooling change on average health.[24] 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. Appendix Figures B1-B4 assess the sensitivity of Figure 3 to changes in the bandwidth and to using linear trends.

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.

The p-value of a test of the difference between the effects on the body size and the blood pressure indices is 0.004. The difference between the effects on the lung function and the blood pressure indices has a p-value of 0.069. The difference between the effects on the body size and the lung function indices has a p-value of 0.906. These results, notably the difference between blood pressure and body size, point to the importance of treating health as multi-

[24] Notice that while Figure 3 uses quarter of birth Table 2 uses day of birth (e.g., September 1, 1957). The same distinction applies to Figure 1 and Table 1.

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.

Our estimates lie within the 95% confidence intervals of Clark and Royer (2010) – see Appendix Table D3 and Appendix Figure D7.[25] Clark and Royer (2010) do not estimate the effects on systolic blood pressure or on lung function. As discussed above, even though we have a larger sample than Clark and Royer (2010), we have a smaller first stage, which explains why our standard errors are not substantially smaller than theirs.

When making this comparison, one should keep in mind that the UK Biobank is not a nationally representative sample. It recruited only individuals living within 25 miles of one of the 22 assessment centers, all of which were located in urban areas. Fry et al. (2017) also document that younger, males living in socioeconomically deprived areas and in particular regions of the UK were less likely to accept the invitation to join the study.

In Appendix H, we compare *compliers* in the UK Biobank to *compliers* in the Health Survey for England – a nationally representative sample – in terms of objectively-measured health outcomes, namely BMI, waist-hip ratio, and diastolic and systolic blood pressures.[26] We find that compliers in the UK Biobank have lower BMIs and waist-hip ratios, but that they also have higher diastolic and systolic blood pressures.[27] Thus, we do not find unequivocal evidence that the UK Biobank sample is healthier than the average population (among compliers). Moreover, while these differences in health are statistically significant, their magnitudes are

[25] In contrast, our results lie outside the confidence intervals of Davies et al. (2017), which study the same reform and data (UK Biobank). We believe this is due to differences in the weighting procedure they use, their specification (bandwidth and polynomial choice) and sample selection (they do not include Wales and Scotland).

[26] We approximate compliers as those born before September 1957 who dropped out of school at age 15 or younger. The UK Biobank sample is restricted to respondents living in England. We focus on objectively-measured health outcomes to avoid concerns that differences in self-reported measures may be partly due to differences in question wording.

[27] Fry et al. (2017) also find that the UK Biobank sample has lower BMI than the Health Survey for England sample. Notice, however, that they are not comparing compliers. They also do not compare the two samples in terms of blood pressure.

small. Finally, for each health outcome, there is substantial overlap between the distribution in the UK Biobank sample and in the Health Survey for England sample.

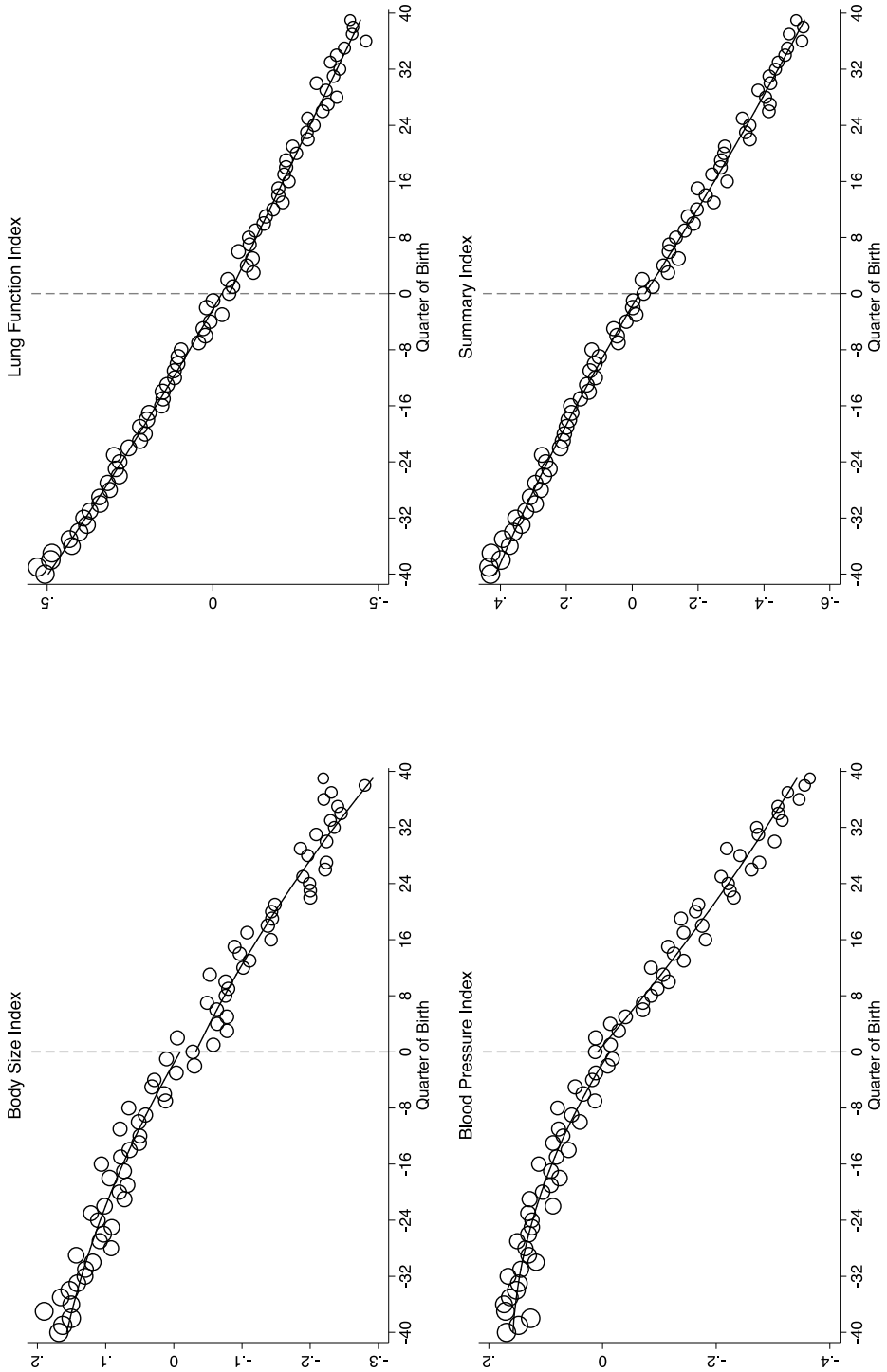
We also used genetic data and a nationally representative sample, the English Longitudinal Survey of Ageing (ELSA), to investigate the selectivity of the UK Biobank sample. Genetic data may be useful in this regard because it is fixed at conception. We calculated the distribution of the polygenic score for educational attainment—an index created from millions of genetic markers—in the ELSA and the UK Biobank samples. Appendix figure H5 shows that, while there are differences, there is substantial overlap in the support of the two distributions. Moreover, despite concerns about sample selection, our results retain their internal validity—see Appendix A.

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.

3. Methods for Distributional Effects Estimates

Even though the effects on average health are informative, they may conceal larger effects on particular parts of the health distribution with important policy implications. Moreover, the average effects documented here are suggestive at best, being only significant at the 10% level. As illustrated in Appendix F, if the effects of education on health are concentrated at particular parts of the health distribution, a distributional test may be better powered than a test of difference in means. In this section, we describe the methods we use to estimate how education affects the distribution of health.

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.[28]

The assumptions of our proposed method are similar to those of a traditional RDD (Imbens and Lemieux 2008). The key difference is that, while a traditional RDD requires that the expectations of the potential outcomes are continuous in the running variable at the threshold, our method requires that the distributions of potential outcomes are continuous in the running variable at that point (Shen and Zheng 2015). While we cannot test this assumption formally, we find no significant discontinuity in the distributions of several predetermined variables, strengthening our confidence that this assumption holds—see Appendix A.

[28] 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.

We want to estimate the local distributional treatment effect (LDTE) for compliers. Let the *pre-reform CDF* be the CDF for compliers in the limit when date of birth is converging to September 1, 1957 from the left (i.e., $DoB < 0$):

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

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

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

The LDTE, which is the discontinuity in the CDF at September 1, 1957 (i.e., $DoB = 0$), is estimated as the difference between $F_{post}(h)$ and $F_{pre}(h)$ at a given h:

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

To estimate $\mu(h)$, in practice we discretize the support of the distribution of health and then estimate (5) for each grid point h:[29]

$$\mathrm{I}(Health_i \leq h) = \theta_0(h) + \theta_1(h) Educ_i + l(DoB_i; h) + \boldsymbol{x}_i' \boldsymbol{\theta}_2(h) + \varepsilon_i(h), \tag{5}$$

[29] For each one of the three health indices, we first calculate the 1st and the 99th percentiles among those born before September 1, 1957:

$$0.01 = \Pr(Body\ Size_i \leq q_1^{bs} | DoB < 0) = \Pr\left(Lung\ Function_i \leq q_1^{lf} | DoB < 0\right) =$$
$$= \Pr\left(Blood\ Pressure_i \leq q_1^{bp} | DoB < 0\right)$$
$$0.99 = \Pr(Body\ Size_i \leq q_{99}^{bs} | DoB < 0) = \Pr\left(Lung\ Function_i \leq q_{99}^{lf} | DoB < 0\right)$$
$$= \Pr\left(Blood\ Pressure_i \leq q_{99}^{bp} | DoB < 0\right)$$

Next, we define the starting point of the grid as the minimum of the 1st percentiles among the health indices and the endpoint as the maximum of the 99th percentiles among the health indices:

$$h_1 = min\{q_1^{bs}, q_1^{lf}, q_1^{bp}\}$$
$$h_{21} = max\{q_{99}^{bs}, q_{99}^{lf}, q_{99}^{bp}\}$$

The grid consists of 21 points uniformly distributed between h_1 and h_{21}.

where $\mathrm{I}(Health_i \leq h)$ is an indicator variable for whether the health index of individual i is smaller or equal to h. The function $l(\cdot\,; h)$ capture birth cohort trends, which are allowed to differ on either side of the cutoff date *and to vary with* h. It is approximated by a quadratic polynomial of date of birth in days.[30] We estimate (5) 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 coefficient on the latter, $\theta_1(h)$, estimates the discontinuity in $\Pr(Health_i \leq h)$ at $DoB_i = 0$ and thus provides an estimate of $\mu(h)$.[31]

We find it easier to visualize $\hat{\mu}(\cdot)$ by plotting in the same graph the pre-reform CDF $F_{pre}(\cdot)$ and the post-reform CDF $F_{post}(\cdot)$.[32] For a given grid point h, we proceed in three steps. First, we estimate $\mu(h)$. Second, we estimate $F_{pre}(h)$ – see next paragraph. Third, we estimate $F_{post}(h)$ as the sum of $\hat{F}_{pre}(h)$ and $\hat{\mu}(h)$. We repeat this procedure for each one of the 21 grid points. Finally, we draw the estimated pre-reform and post-reform CDFs by drawing $\hat{F}_{pre}(h)$ and $\hat{F}_{post}(h)$ against h. For any given h, the vertical distance between $\hat{F}_{post}(h)$ and $\hat{F}_{pre}(h)$ is equal to $\hat{\mu}(h)$.

To estimate the pre-reform CDF $F_{pre}(\cdot)$, we restrict the sample to respondents born before September 1, 1957 and who left school at age 15 or younger (i.e., "the compliers") and estimate equation (6) for each grid point h:

$$\mathrm{I}(Health_i \leq h) = \delta_0(h) + k(DoB_i; h) + \xi_i(h), \qquad (6)$$

[30] In our main specification with a quadratic polynomial:

$$l(DoB_i; h) = \lambda_1(h)DoB_i + \lambda_2(h)[DoB_i]^2 + \lambda_3(h)I\{DoB_i \geq 0\}DoB_i + \lambda_4(h)I\{DoB_i \geq 0\}[DoB_i]^2.$$

[31] Consider (7), which is the reduced-form version of (5). There is an RD graph for each grid point h for a total of 21 graphs (per health index). In Appendix Figures B11-B13, we show corresponding graphs for the 10th, 25th, 50th, 75th, 90th, and 95th percentiles of the distribution of those born between September 1, 1956 and August 31, 1957.

[32] In the bottom panel of Appendix Figures B14-B16 we plot $\hat{\mu}(h)$ against h.

where the function $k(\cdot\,;h)$ captures *pre-reform* birth cohort trends in health.[33] We can closely represent compliers born before September 1, 1957 by making this sample restriction because 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). The coefficient on the constant, $\delta_0(h)$, estimates $\Pr(Health_i \leq h)$ as $DoB \to 0^-$ and thus provides an estimate of $F_{pre}(h)$. Note that we estimate $F_{pre}(h)$ in order to graphically display our results, but none of our findings depends on $F_{pre}(h)$ (e.g. results in Table 3 do not depend on estimates of $F_{pre}(h)$).

Inference based on the standard errors generated by 2SLS estimates of (5) 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:

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

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).[34] In contrast to Shen and Zhang (2016), which implements a Kolmogorov-Smirnov test, we perform an Anderson-

[33] In our main specification with a quadratic polynomial: $k(DoB_i; h) = \phi_1(h)DoB_i + \phi_2(h)[DoB_i]^2$. These terms take into account any secular trends in our outcomes.
[34] 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.

Darling test (Anderson and Darling 1952) using the following weighted integral as our test statistic:[35]

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

Average treatment effects may not be well-powered to detect effects of education on health that are concentrated on the tails of the health distribution (see Appendix F). We chose the Anderson-Darling test because it is uniformly powered for the whole range $\tau \in [0, 1]$ (Stephens 1974). In contrast, the Kolmogorov-Smirnov test is better powered to detect deviations of the distribution near the median. To test for differences in the bottom 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 top half by integrating from 0.5 to 1.[36]

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.

The CDF approach described above is closely related to a quantile IV approach. The CDF approach is based on the vertical distance between the pre- and post-reform CDFs

[35] 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).

[36] 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.

whereas a quantile approach is based on the horizontal distance between these two CDFs. Therefore, either approach would 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. Nevertheless, when we present our results, we also discuss the effects on some particular quantiles of interest.

It is important to note that our method only recovers differences in the health distribution of compliers and it does not rely on rank preservation assumptions. Stronger assumptions would be needed to recover the distribution of treatment effects, but this is beyond the scope of our paper. Although we do not estimate the distribution of treatment effects, we can examine whether our estimates of the effect on the health distribution are consistent with homogeneous treatment effects. We show in Appendix I that, if treatment effects were homogeneous, our estimates of distributional effects would behave predictably. Namely, we would see the largest distributional effects where the largest mass of compliers is: around the mode of the distribution of compliers. Therefore, any other pattern of results is indirect evidence of heterogeneous treatment effects.

Importantly, our two-stage least squares estimates of distributional effects are not likely to be biased by heterogeneity in the first stage. There are two potential forms of first-state heterogeneity for distributional effects: extensive- and intensive-margin heterogeneity. Extensive-margin heterogeneity means that the fraction of individuals staying in school for longer as a result of the 1972 ROSLA may vary depending on where in the health distribution these individuals would have been absent the reform. For example, unhealthy individuals may be more likely to be drop out of school at age 15 before the reform relative to their healthy peers and therefore may be more likely to be "compliers" in this setting. Intensive-margin heterogeneity means that, among compliers, one's additional schooling associated with the ROSLA may vary depending on where in the health distribution s/he would have been in the

absence of the reform. For example, healthy compliers may be more likely to stay in school for greater than one year due to the ROSLA despite only being compelled to stay for one.

Under the assumptions of the regression discontinuity design, extensive-margin heterogeneity does not affect on our two-stage least squares estimates (see Appendix J). The main intuition behind the proof is that two-stage least squares estimates of distributional effects recover differences in the health distribution of only compliers, who, by definition, were all affected by the policy. However, if there is heterogeneity on the intensive margin, the estimates of the distributional effects will overweigh the treatment effects of those who stayed in school longer. As a result, the largest distributional effects would not be observed at the mode of the distribution of compliers; instead, the estimates would be biased toward the mode of distribution of those who increased their schooling the most as a result of the ROSLA.

In practice, it is unlikely that heterogeneity on the intensive margin could explain our results. The main effect of the ROSLA was to keep in school until age 16 students who would have dropped out by age 15 otherwise; the increase in the fraction of students staying in school until at least age 17 as a result of the ROSLA is small: about 2 percentage points (compared to 15 percentage points at age 16, see Appendix K). Clark and Royer (2013) reach a similar conclusion: "the 1972 change had small, at best, effects on the fractions completing 11 or fewer years…To a first approximation therefore, one can view these law changes as forcing students that would previously have left at the earliest opportunity to stay in school for one more year" (pg. 2102). These suggest little scope for first-stage, intensive-margin heterogeneity in our context.

4. Distributional Effects of Education on Health

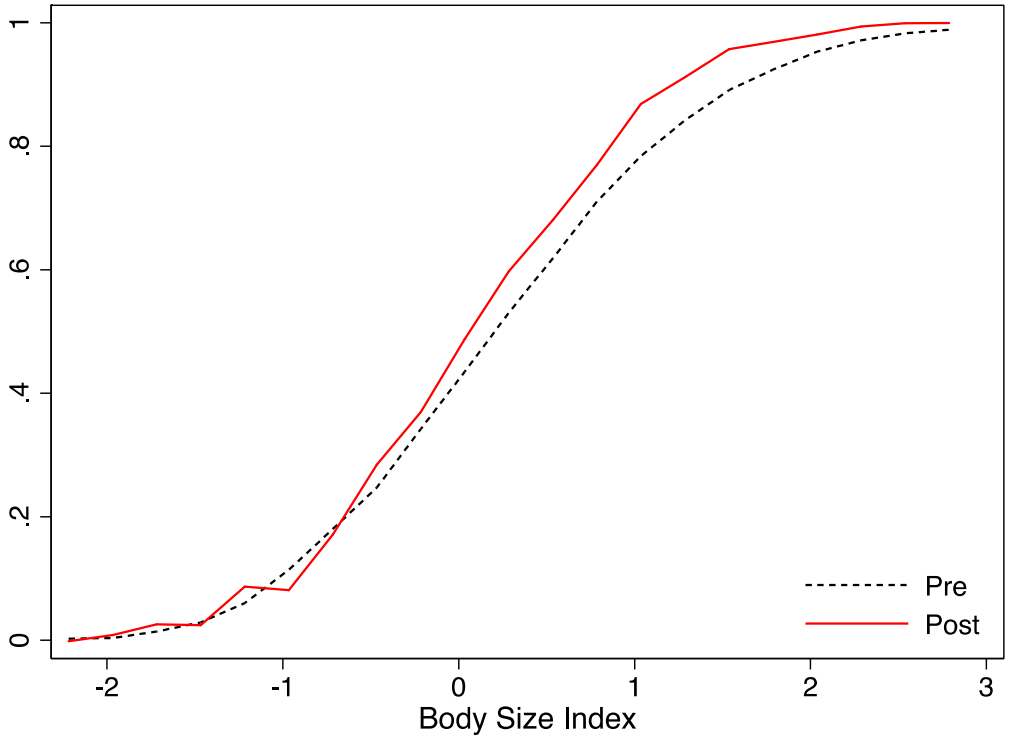
Figure 4 shows the distributional treatment effects of education on body size. It shows the pre- and post-reform CDFs of the body size index for compliers. As explained in Section

3, the pre-reform CDF is obtained by estimating (6) for each grid point h and then plotting $\hat{\delta}_0(h)$ against h.[37] The discontinuity in the CDF, $\mu(h)$, is obtained by estimating (5) through 2SLS for each grid point h. The post-reform CDF at a given grid point h is obtained by adding $\hat{\mu}(h)$ "vertically" to $\hat{\delta}_0(h)$.

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. This result is consistent with a Grossman-type model with diminishing marginal utility of health (Kippersluis and Galama 2014). 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.

[37] 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.

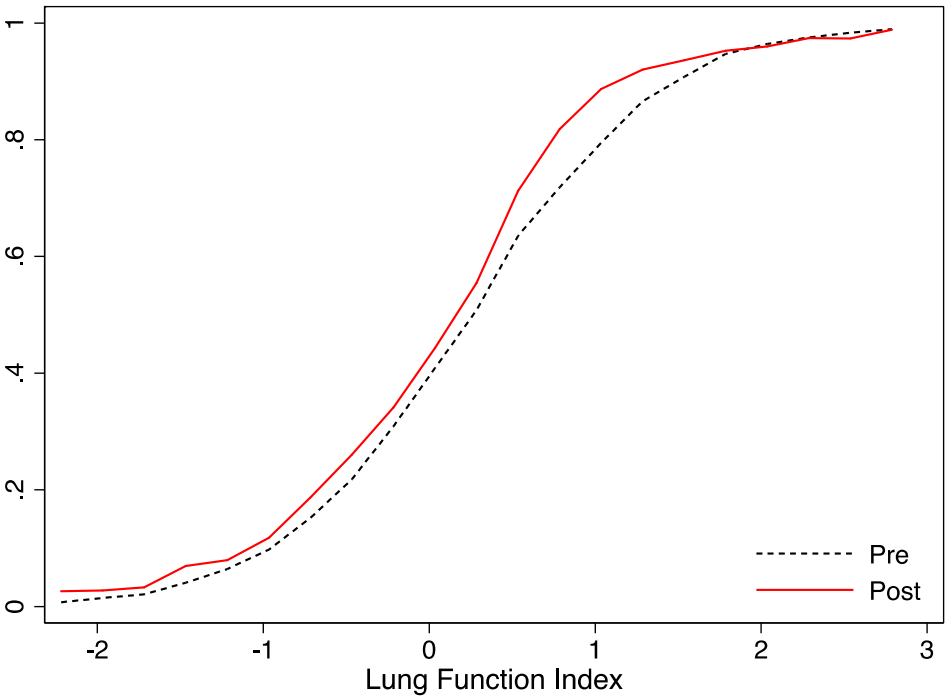
Figure 4: Distributional Effects on Body Size Index



Notes: The figure shows the pre- and post-reform CDFs of the body size index for compliers. N = 266,525.

While Figure 4 has the advantage of graphically displaying the distributional treatment effects in just one figure, it lacks the transparency of traditional RDD graphs. In Appendix Figure B11, we present traditional RDD graphs for the 10th, 25th, 50th, 75th, 90th, and 95th percentiles of the distribution of the body size index (of those born between September 1, 1956 and August 31, 1957). Appendix Figures B12 and B13 present corresponding graphs for the lung function and blood pressure indices.

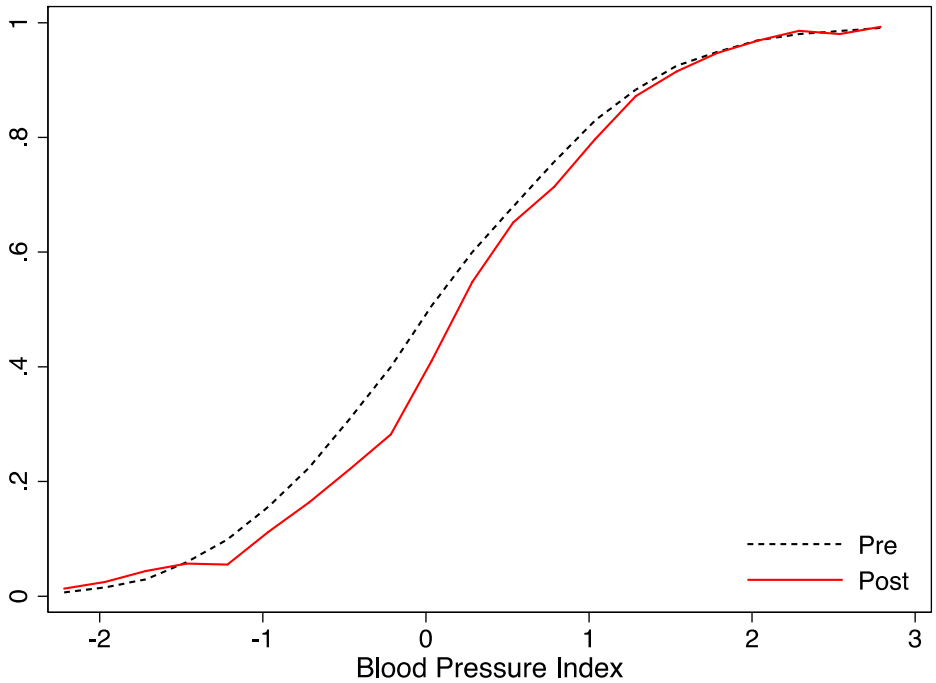
Figure 5: Distributional Effects on Lung Function Index



Notes: The figure shows the pre- and post-reform CDFs of the lung function index for compliers. N = 215,366.

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 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.

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.[38] 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 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).

[38] 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.

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. Appendix Figures B5-B10 assess the sensitivity of Figures 5-7 to changing the bandwidth, using linear trends, and the inclusion of controls.

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.

We can reject the null for the top half of the body size distribution and for the bottom half of the blood pressure distribution (at the 5% significance level). The p-value for the top half of the lung function distribution is 0.0618. Appendix Tables B1-B3 assess the sensitivity of these results to changing the bandwidth, using linear trends, and including controls. While the p-values change across specifications, the main patterns remain: in most cases, we can reject the null of equality for the top half of the body size distribution and for the bottom half of the blood pressure distribution.

Table 3: P-values of Distributional Tests

	Body Size	*Lung Function*	*Blood Pressure*
Full Distribution	0.0896	0.1712	0.0362
Bottom Half	0.9526	0.5962	0.0102
Top Half	0.0126	0.0618	0.1502

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.

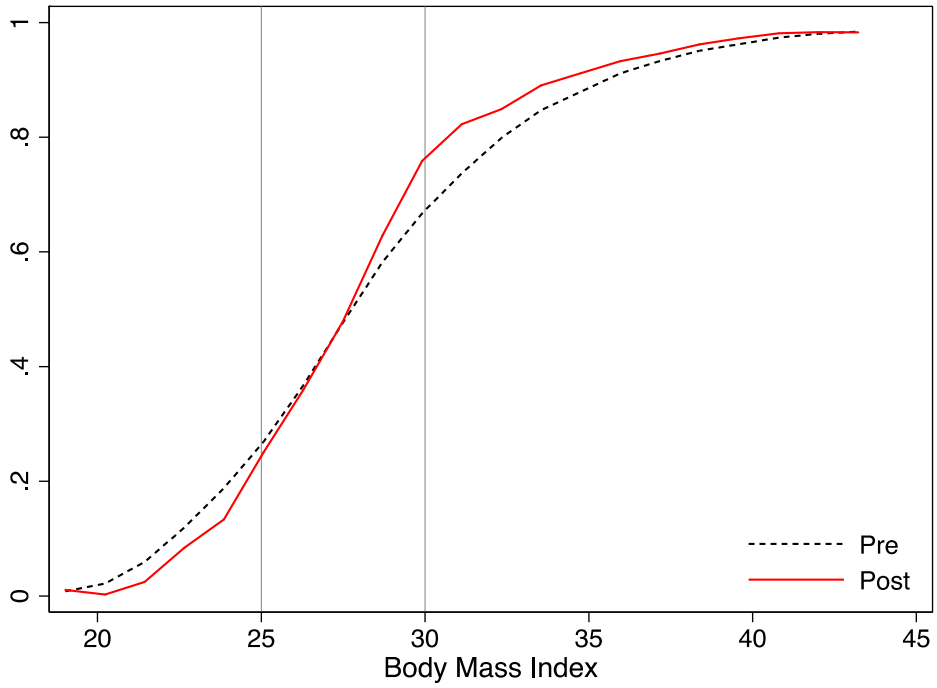
Figures 7 and 8 shed light on these findings by plotting results for measures with clinical thresholds. Figure 7 shows the pre- and post-reform CDFs of body mass index (for compliers). Figure 8 shows the pre- and post-reform CDFs of diastolic blood pressure (for compliers).[39]

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 stage 2 hypertension (classified as having a diastolic blood pressure above 90 mmHg). Staying in school until age 16 *increases* the probability of stage 1 hypertension (defined as having diastolic blood pressure between 80 mmHg and 90 mmHg; known as "prehypertension" before the 2017 redefinition) by 7.9 percentage points.[40]

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

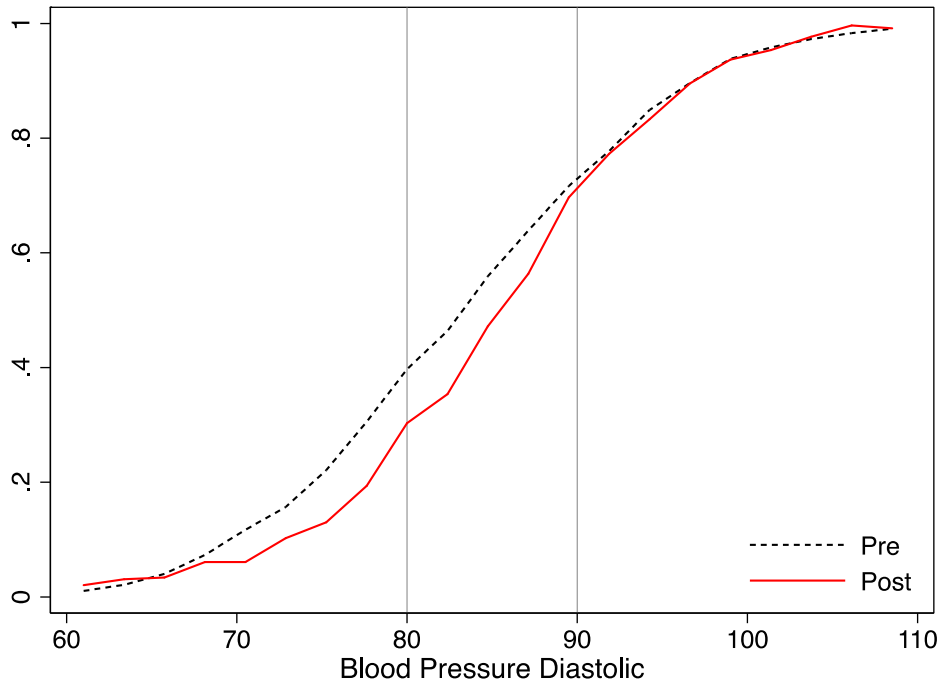
[40] Despite no change in stage 2 hypertension, the increase in the fraction of people in the stage 1 (formerly pre-high) range might still mean a worsening of health, as observational studies indicate that death from both ischemic heart disease (IHD) and stroke increases progressively and linearly from 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). However, stage 2 hypertension is more serious than stage 1: while medications are the main treatment for stage 2 hypertension, for stage 1 the focus is on lifestyle changes (Bakris and Sorrentino 2018).

Figure 7: Distributional Effects on Body Mass Index



Notes: The figure shows the pre- and post-reform CDFs of body mass index for compliers. N = 270,019.

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.

These results illustrate the importance of studying distributional effects. While the average treatment effects show improvements in body size 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 observed consequences for the prevalence of stage 2 hypertension. 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.

Our findings are not consistent with homogeneous treatment effects of education on health. The main effect of the schooling reform was to induce students who would have dropped out at age 15 (i.e., the compliers) to stay in school an additional year. Homogeneous treatment effects do not imply a constant $\theta_1(h)$ for all h in equation (5), or equivalently, a parallel vertical shift in Figures 4-8. Rather, if this additional year of schooling affected the health of all compliers equally, we would see effects across the whole support of the complier distribution, with the largest effect at the mode of the pre-reform distribution. This is because, when more people are near some threshold h_0 absent the reform, more people will cross the h_0 threshold after the reform is in place. Appendix I provides a formal proof.

The results show a different pattern. Figures 4-8 indicate that our effects are limited to some parts of the distributions of health of compliers. Indeed, the statistical tests in Table 3 cannot reject a null effect for the bottom half of the body size and lung function distributions and the top half of the blood pressure distribution. Importantly, the effects we do find are *not* concentrated around the modes of these distributions. For body size, they are concentrated around the 78th percentile of the complier distribution (72nd percentile for lung function and 40th for blood pressure). These patterns imply heterogeneous effects of education on health.

5. Channels

One of the channels through which education may affect health is 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).

Taking advantage of the richness of the UK Biobank data, we investigate whether education has a causal effect 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).[41] Accelerometers worn for 7 days were used to measure physical activity.[42] Smoking was self-reported.

Appendix Table E4 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). There are, however, no effects on total caloric intake, sugars, or carbohydrates. Appendix Table E5 shows no effects on the measures of smoking and physical activity we have available.[43]

This analysis suggests that improved diet is an important channel through which education reduces body size. Those who stayed in school until age 16 had better diets in middle age – about 10% lower in fat and 15% lower in saturated fat. Even if the energy content of

[41] 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.

[42] 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

[43] These are consistent with Clark and Royer (2013), who also find no effects of the 1972 ROSLA on self-reported smoking or physical activity.

one's diet is held constant, changes in diet composition can affect body weight (Hall et al. 2012).

The pathways are less clear for the harmful effect of education on blood pressure. We find no effects on a (self-reported) measure of hypertension diagnosis[44] and on (current) blood pressure medication (Appendix Table E5). 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.[45] 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.

[44] The wording of the question was "Have you been told by a doctor that you have high blood pressure"?
[45] Appendix Table E2 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.

6. Conclusions

In this paper, we investigate how education affects the distribution of health along three dimensions: body size, lung function, and blood pressure. While the main take-away from Clark and Royer (2013) is that education does not affect health, our results suggest a more nuanced story. The distributional analysis shows that education does affect the health of some subpopulations. We find, for example, that the school-leaving age reform shifted part of the body size distribution to the left—i.e., it caused reductions in body size. Importantly, our results do not rely on rank preservation assumptions and or on extensive-margin homogeneity in the effects of the ROSLA on education (i.e., the first stage).

Because the effects of the reform are concentrated at the upper tail of the distribution rather than at the mode, this is evidence that the effect of education on body size is heterogeneous. Indeed, in related work we show that the effect varies with one's genetic predisposition to obesity: the reductions in body size were larger for individuals at greater genetic risk of obesity (Barcellos, Carvalho, and Turley 2018). These findings support Galama et al. (2018)'s hypothesis that heterogeneity in the effects of education on health explain why some studies find that more education causes better health while others find no effect.

We conduct a back-of-envelope calculation to illustrate the importance of accounting for the heterogeneity in the effects of education when evaluating its estimated impacts (Heckman et al. 2016). Take, for example, the relationship between BMI and mortality. It is U-shaped (Fogel 1994; Aune et al. 2016), suggesting that the reduction in mortality caused by a given reduction in BMI is larger for someone obese than for someone with normal weight. A reduction in average BMI of 0.4 caused by staying in school until age 16 (i.e., the average treatment effect on the treated) implies a reduction in the relative risk of death of 0.85%—see Appendix G for details. In contrast, the local distributional treatment effects yield a much larger

reduction of 3.27%. That is because of the U-shaped relationship between BMI and mortality and because the effects on BMI were concentrated at the upper tail of the BMI distribution.

The effects of education on health may vary not only across people but also across different health dimensions. Besides reducing body size, the school-leaving age reform increased blood pressure. Education seems to have reduced body size through improvements in SES and diet, but it is not clear the channels through which education increased blood pressure. Interestingly, the improvement in body size was concentrated at the top half of the body size distribution (among the least healthy) while the worsening in blood pressure was concentrated at the bottom half of the blood pressure distribution (among the most healthy).

These results underscore 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 BMI occurred at the upper tail of the BMI distribution, staying in school until age 16 reduced obesity rates (i.e., BMI above 30) by 7.5 percentage points. In contrast, the increase in blood pressure was concentrated below the clinical threshold for stage 2 hypertension (i.e., diastolic blood pressure above 90 mmHg or systolic blood pressure above 140) with no statistically significant consequences for prevalence.

REFERENCES

Al Snih, S., Ottenbacher, K.J., Markides, K.S., Kuo, Y.F., Eschbach, K. and Goodwin, J.S., 2007. The effect of obesity on disability vs mortality in older Americans. Arch Intern Med, 167(8):774-80

N Allen et al. 2012: UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology Volume 1, Issue 3:123-126.

Albouy, Valerie and Laurent Lequien. 2009. “Does compulsory education lower mortality?” *Journal of Health Economics*, 28,155–168. Journal of Health Economics, Volume 28, Issue 1:155-168

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, T. W.; Darling, D. A. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. Ann. Math. Statist. 23 (1952), no. 2, 193--212. doi:10.1214/aoms/1177729437. https://projecteuclid.org/euclid.aoms/1177729437

T. W. Anderson & D. A. Darling (1954) A Test of Goodness of Fit, Journal of the American Statistical Association, 49:268, 765-769, DOI: 10.1080/01621459.1954.10501232

Aune, D., Sen, A., Prasad, M., Norat, T., Janszky, I., Tonstad, S., Romundstad, P. and Vatten, L.J.. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. BMJ. 2016 May 4;353:i2156. doi: 10.1136/bmj.i2156.

Avendano, Mauricio, Augustin de Coulon and Vahe Naflyan. 2017. “Does more Education always improve mental health? Evidence from a British compulsory schooling reform.” Working paper, University of York.

Bakris, G., & Sorrentino, M. 2018. Redefining Hypertension—Assessing the New Blood-Pressure Guidelines. *New England Journal of Medicine*. N Engl J Med. 2018 Feb 8;378(6):497-499. doi: 10.1056/NEJMp1716193.

Barcellos, Silvia Helena, Leandro Carvalho, and Patrick Turley. 2018. “Education can Reduce Health Differences Related to Genetic Risk of Obesity.” *Proceedings of the National Academy of Science*, 115(42): E9765-E9772. https://doi.org/10.1073/pnas.1802909115

Becker, G.S., Mulligan, C.B. Becker, G., & Mulligan, C. 1997. The Endogenous Determination of Time Preference. The Quarterly Journal of Economics, 112(3), 729-758. Retrieved from http://www.jstor.org/stable/2951254

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.

Marianne P. Bitler & Hilary W. Hoynes & Thurston Domina, 2014. "Experimental Evidence on Distributional Effects of Head Start," NBER Working Papers 20434, National Bureau of Economic Research, Inc.

Black, Dan, A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel. 2003. "Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System." American Economic Review, 93 (4): 1313-1327. DOI: 10.1257/000282803769206313

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(7), 1285-93. PMCID: PMC3713113

Chan, June M., Eric B. Rimm, Graham A. Colditz, Meir J. Stampfer, and Walter C. Willett. "Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men." Diabetes Care. 1994 Sep;17(9):961-9.

Chen, E., Miller, G. E., Brody, G. H., & Lei, M. (2014). Neighborhood Poverty, College Attendance, and Diverging Profiles of Substance Use and Allostatic Load in Rural African American Youth. Clinical psychological science : a journal of the Association for Psychological Science, 3(5), 675-685. PMCID: PMC4565724

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), 2560-2572. DOI: 10.1001/jama.289.19.2560

Clark D,Royer H. 2010. "The Effect of Education on Adult Health and Mortality: Evidence from Britain." NBER Working Paper No. 16013. DOI: 10.3386/w16013

Clark D Royer H. 2013. "The Effect of Education on Adult Mortality and Health: Evidence from Britain." Am Econ Rev. 103(6):2087-120.

Conti, G., Heckman, J., and Urzua, S. 2010. "The education-health gradient." Am Econ Rev., 100(2): 234–238. doi: 10.1257/aer.100.2.234. PMCID: PMC3985402

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

Economic Review: Papers & Proceedings 100 (May 2010): 234–238. http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.234

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. Occup Environ Med.,73(6):378-84.

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. https://doi.org/10.1111/j.1468-0262.2007.00738.x

Fogel, R.W., 1994. Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy. *American Economic Review*, 84(3),369-395.

Galama, Titus J., Adriana Lleras-Muney, and Hans van Kippersluis. The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence. No. w24225. National Bureau of Economic Research, 2018. Oxford Research Encyclopedia of Economics and Finance, forthcoming.

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. Br J Nutr. 2016 Feb 28;115(4):681-6. doi: 10.1017/S0007114515004821

Gelman, Andrew and Guido Imbens. 2014. "Why high-order polynomials should not be used in regression discontinuity designs." Unpublished Manuscript. Published: Andrew Gelman & Guido Imbens (2017) Why high-order polynomials should not be used in regression discontinuity designs, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2017.1366909

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. https://doi.org/10.1111/j.1467-9442.2012.01739.x

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. Proc Natl Acad Sci U S A., 6;99(16):10929-34. PMCID:PMC125075 DOI:10.1073/pnas.162086599

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 WeightRegulation—Consensus Statement from the American Society for Nutrition.” Am J Clin Nutr, 95(4): 989–994. doi: 10.3945/ajcn.112.036350.

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. *Am J Epidemiol*, 122(2), 341-344.

Janke, K., Johnston, D.W., Propper, C. and Shields, M.A. 2018. “The Causal Effect of Education on Chronic Health Conditions.” Unpublished manuscript. IZA DP No. 11353. https://www.iza.org/publications/dp/11353/the-causal-effect-of-education-on-chronic-health-conditions

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. DOI:10.1016/j.jhealeco.2012.05.004

Johnston, D.W., C. Propper and M.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. DOI:10.1016/j.jhealeco.2009.02.010

Jürges, Hendrik, Eberhard Kruk and Steffen Reinhold. 2013. “The effect of compulsory schooling on health—evidence from biomarkers.” *J Popul Econ*, 26:645–672. https://doi.org/10.1007/s00148-012-0409-9

Jensen, R. and Lleras-Muney, A., 2012. “Does staying in school (and not working) prevent teen smoking and drinking?” *J Health Econ*, 31(4): 644-657. doi: 10.1016/j.jhealeco.2012.05.004

Kannel, William B. 1996. "Blood pressure as a cardiovascular risk factor: prevention and treatment." JAMA 275(20) .: 1571-1576.

Kemptner, Daniel, Hendrik Jürgesb and Steffen Reinholdc. 2011. “Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany.” J Health Econ. 30, 340–354. DOI:10.1016/j.jhealeco.2011.01.004Lee, D. S. 2008. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142(2):675 – 697.

Kippersluis, Hans van and Titus J. Galama. 2014. “Wealth and Health Behavior: Testing the Concept of a Health Cost.” *European economic review*, 72:197-220.

Kippersluis, Hans van and Titus J. Galama. , Adriana Lleras-Muney, and. The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence. No. w24225. National Bureau of Economic Research, 2018. Oxford Research Encyclopedia of Economics and Finance, forthcoming.

Lee, D.S. and Card, D. 2008. “Regression Discontinuity Inference with Specification Error.” *Journal of Econometrics*, *142*(2): 655-674.

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. 2002. 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*; ;360(9349):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.

Månsson, N.O., Eriksson, K.F., Israelsson, B., Ranstam, J., Melander, A. and Råstam, L., 1996. Body mass index and disability pension in middle-aged men—non-linear relations. *Int J Epidemiol*, 25(1),80-5.

Meghir, C., M. Palme and E. Simeonova, f "Education, Health and Mortality: Evidence from a Social Experiment", American Economic Journal: Applied Economics. 10(2). DOI: 10.1257/app.20150365

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. *Proc Natl Acad Sci U S A., 112(33),* 10325–30. PMCID:PMC4547243. DOI: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?" *J Health Econ*, 27(6), 1516-31. DOI:10.1016/j.jhealeco.2008.07.005

Perez-Arce, Francisco. 2017. "The effect of education on time preferences." *Economics of Education Review.* 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. 1974. "EDF statistics for goodness of fit and some comparisons." *Journal of the American statistical Association.* 69(347): 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

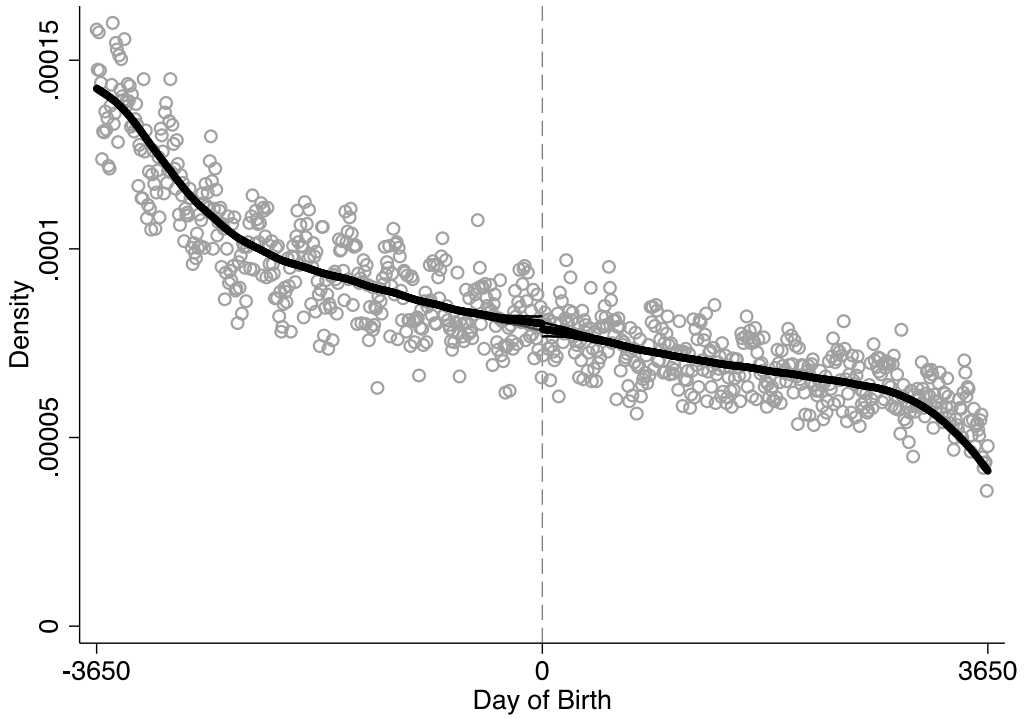
Thomas, P. S., Cowen, E. R., Hulands, G., & Milledge, J. S. 1989. Respiratory function in the morbidly obese before and after weight loss. *Thorax*, *44*(5), 382-386. PMCID: PMC461837

Tockman, Melvyn S., Nicholas R. Anthonisen, Elizabeth C. Wright, and Michele G. Donithan. 1987. "Airways obstruction and the risk for lung cancer." Annals of internal medicine 106(4) 512-518.

Vasan RS, Larson MG, Leip EP, Evans JC, O’Donnell CJ, Kannel WB, et al. Impact of high-normal blood pressure on the risk of cardiovascular disease. *N Engl J Med* 2001;345:1291-7. DOI:10.1056/NEJMoa003417

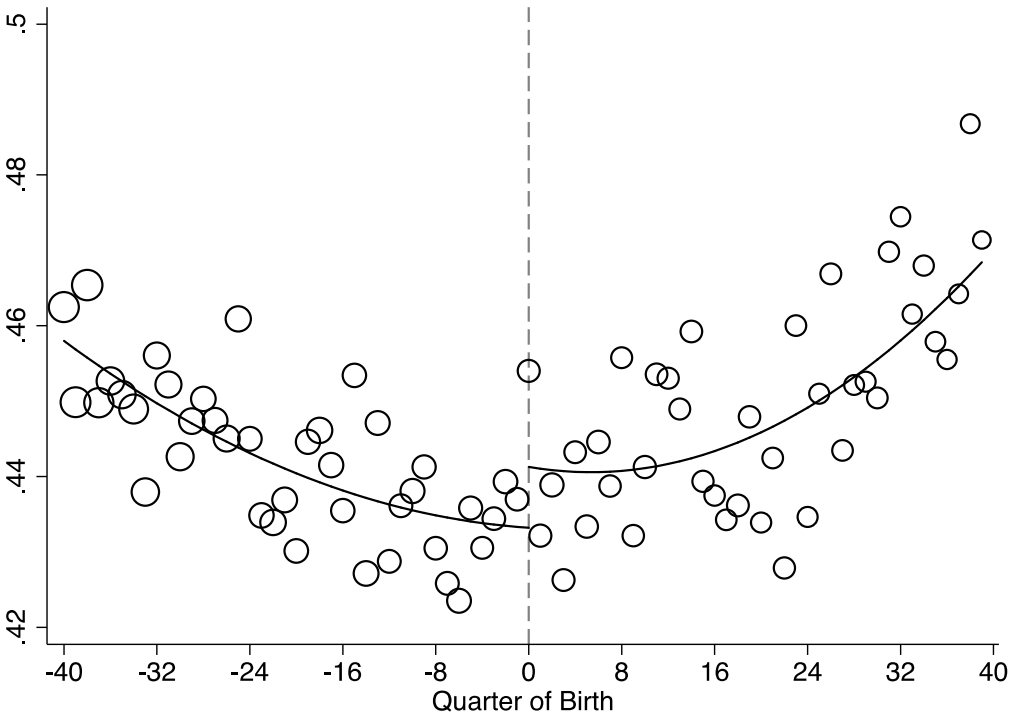
Appendix A

Appendix Figure A1: McCrary Test



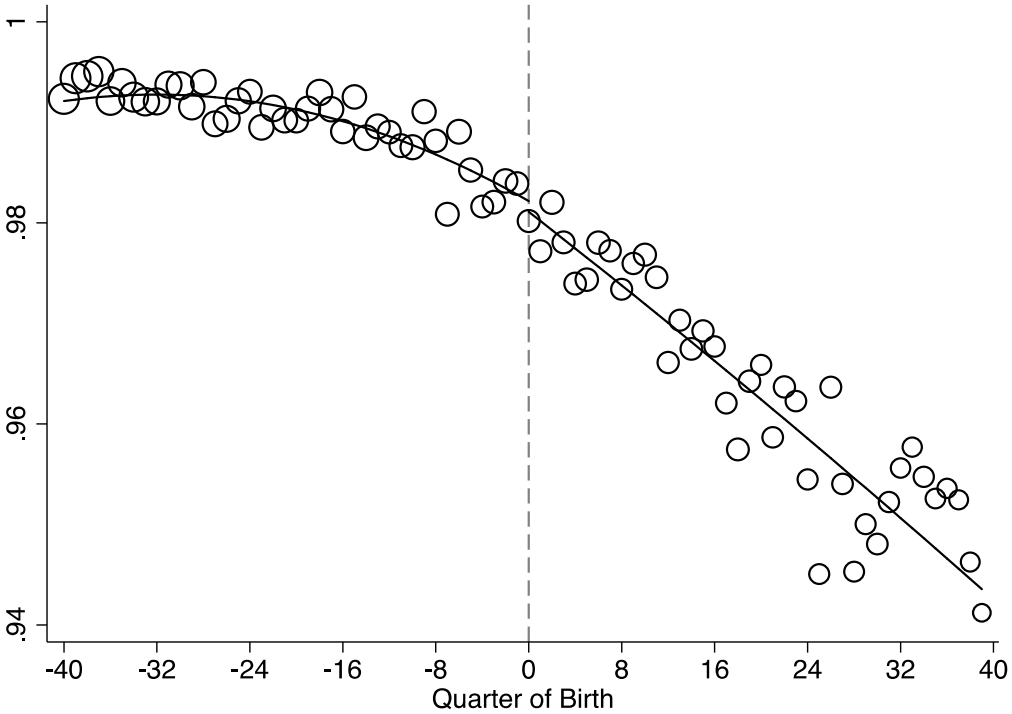
Notes: The figure shows the fraction of study participants by day of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The 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



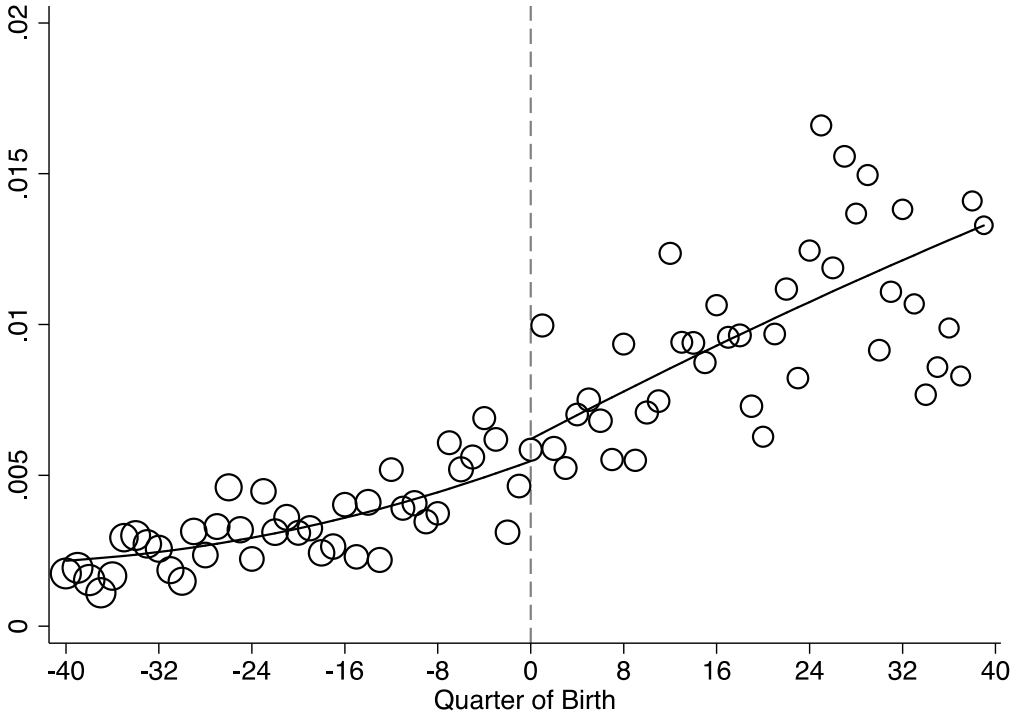
Notes:
Notes:
The figure shows the fraction of male study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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



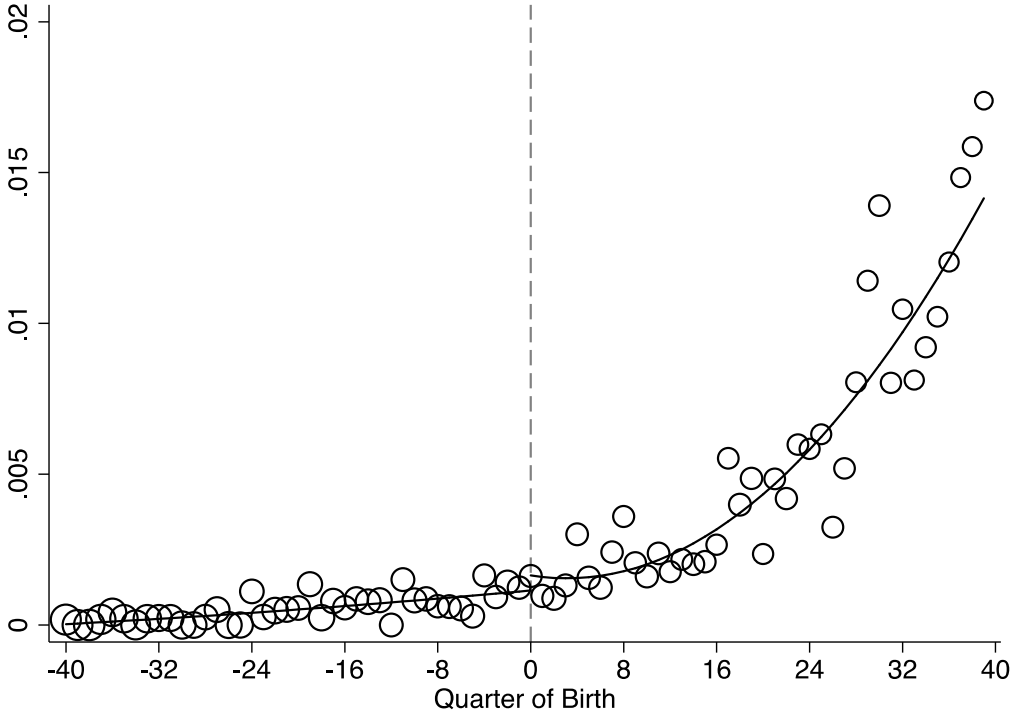
Notes:
The figure shows the fraction of white study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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



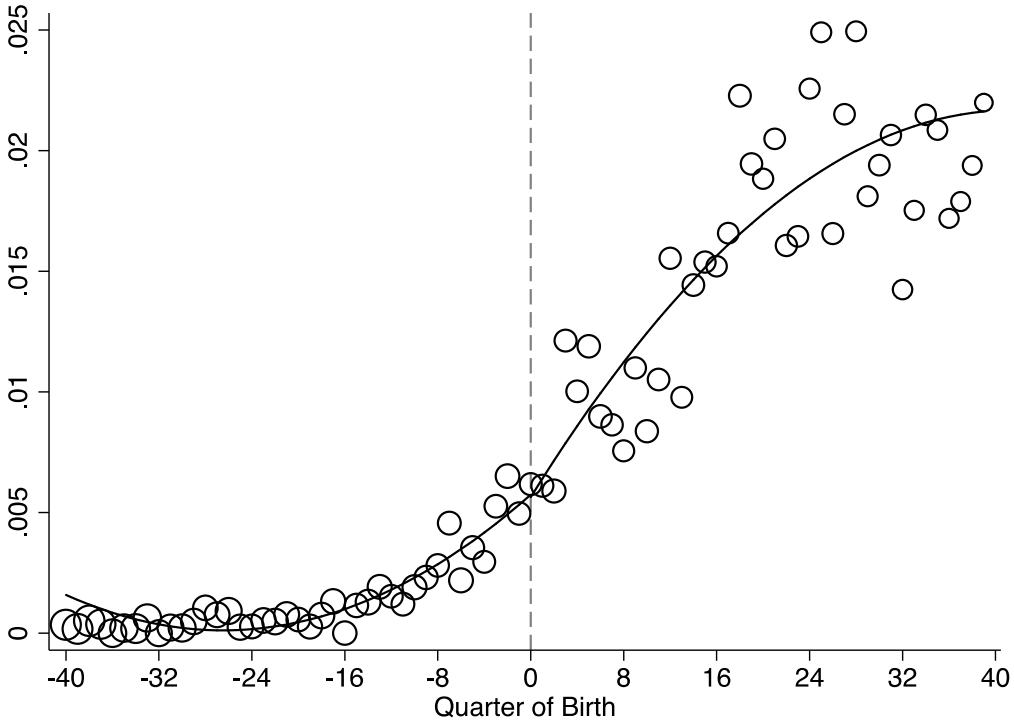
Notes: The figure shows the fraction of study participants of mixed ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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



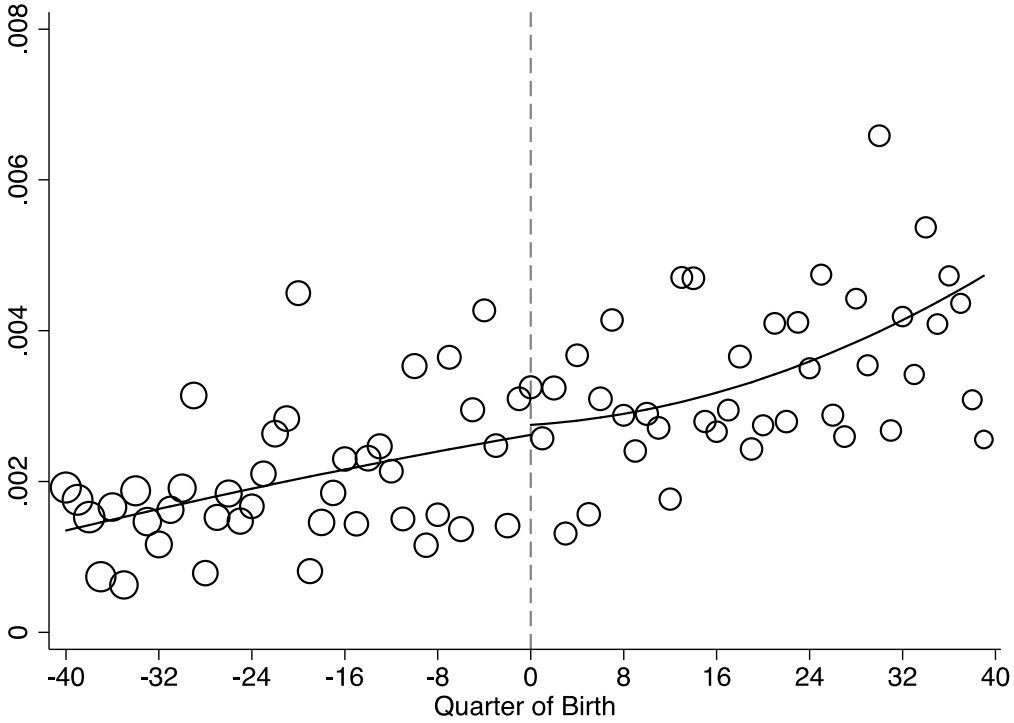
Notes: The figure shows the fraction of Asian study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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



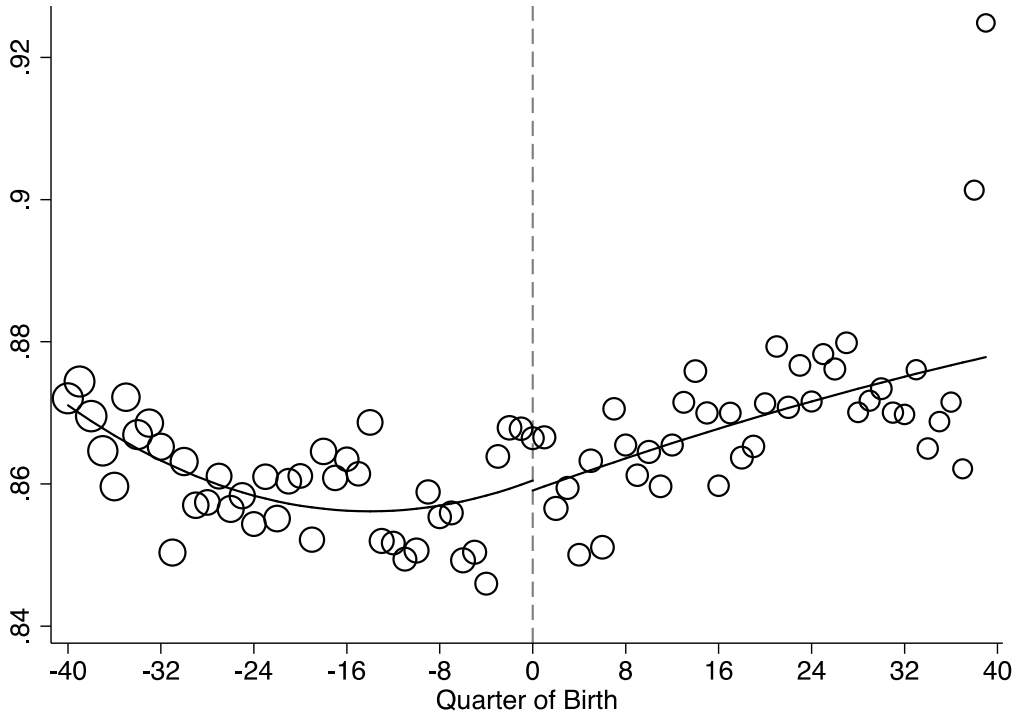
Notes: The figure shows the fraction of black study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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



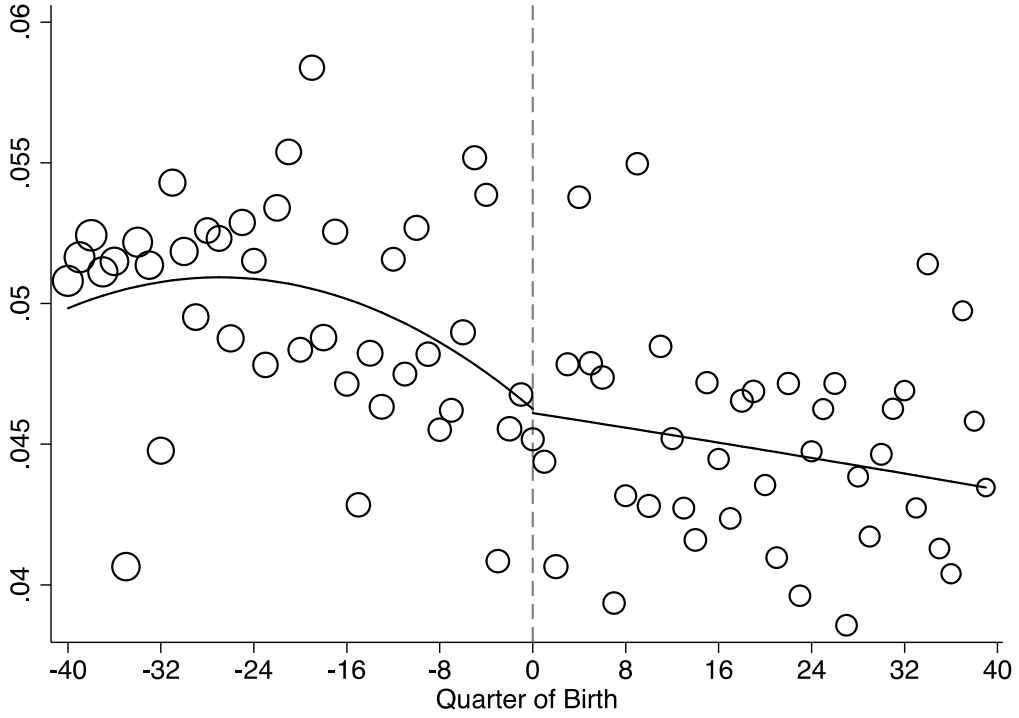
Notes: The figure shows the fraction of study participants of another ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before 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



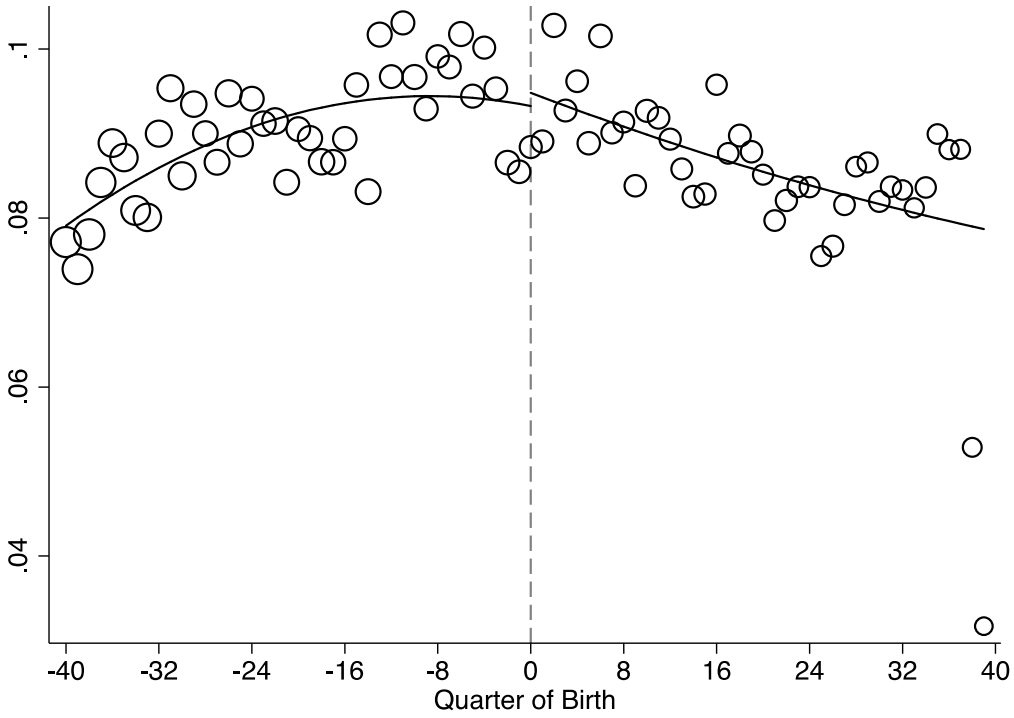
Notes: The figure shows the fraction of study participants born in England by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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



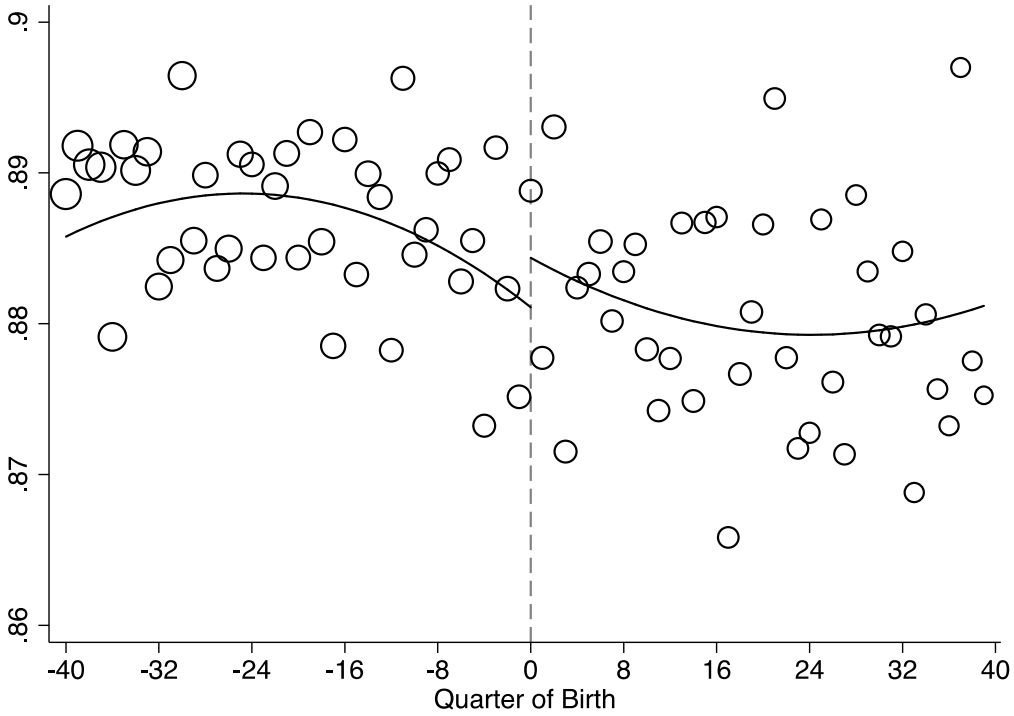
Notes: The figure shows the fraction of study participants born in Wales by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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



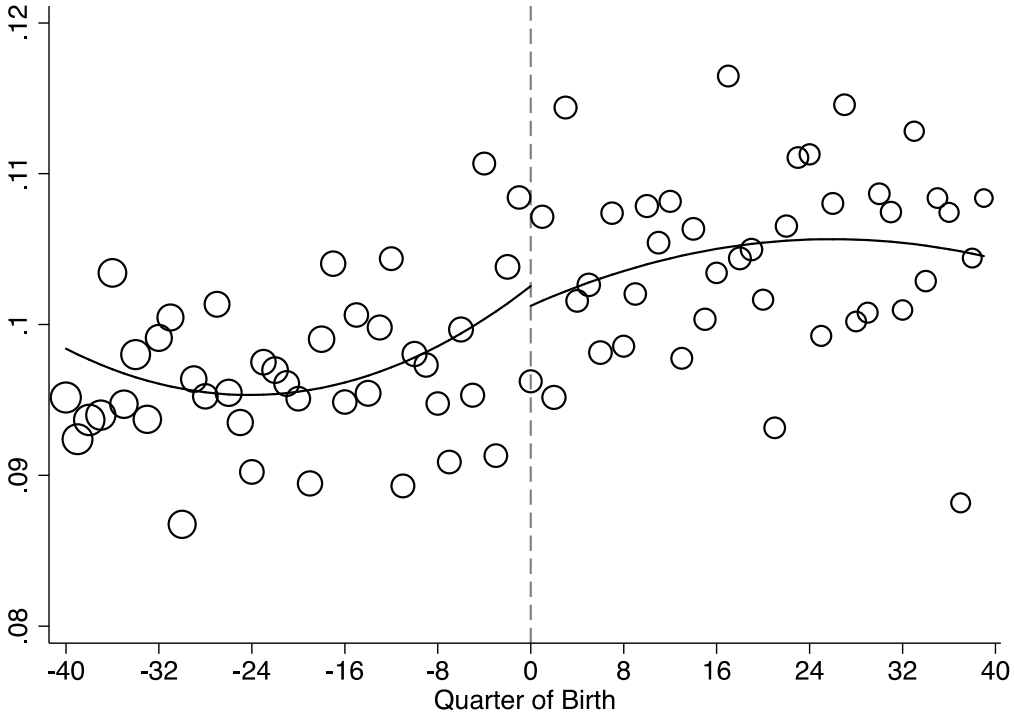
Notes: The figure shows the fraction of study participants born in Scotland by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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



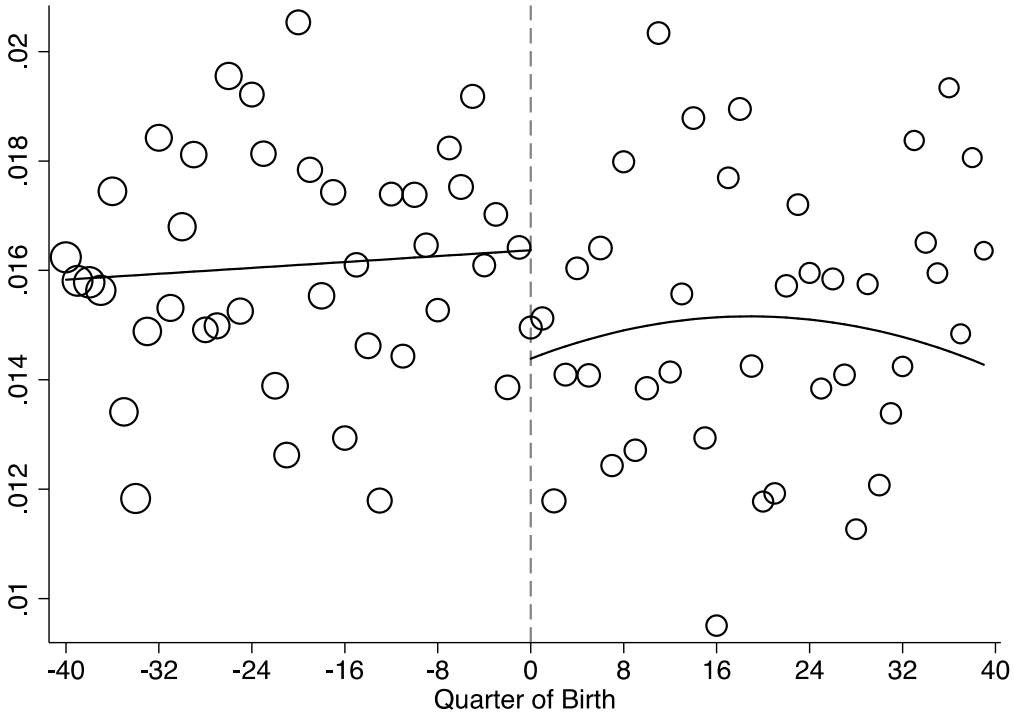
Notes: The figure shows the fraction of right-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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



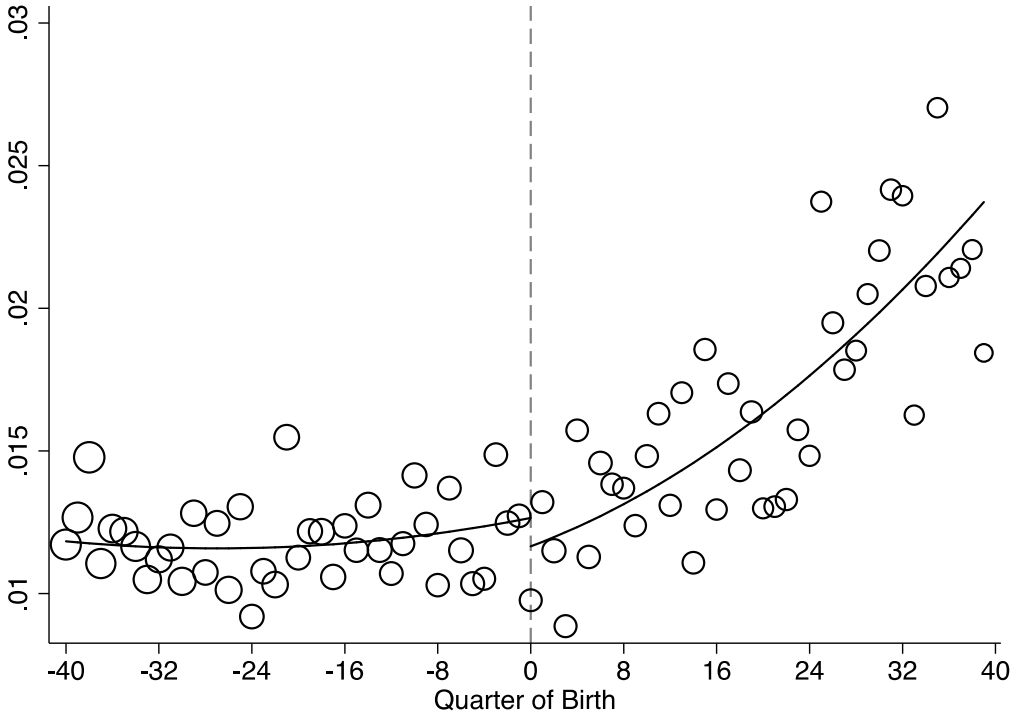
Notes: The figure shows the fraction of left-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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



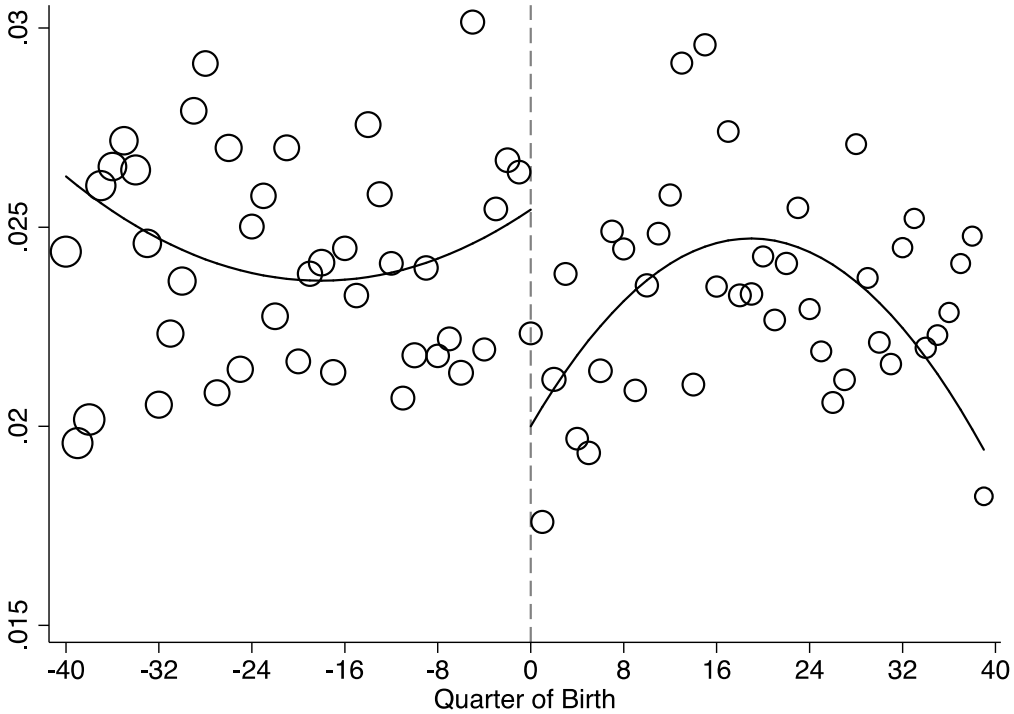
Notes: The figure shows the fraction of ambidextrous study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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.

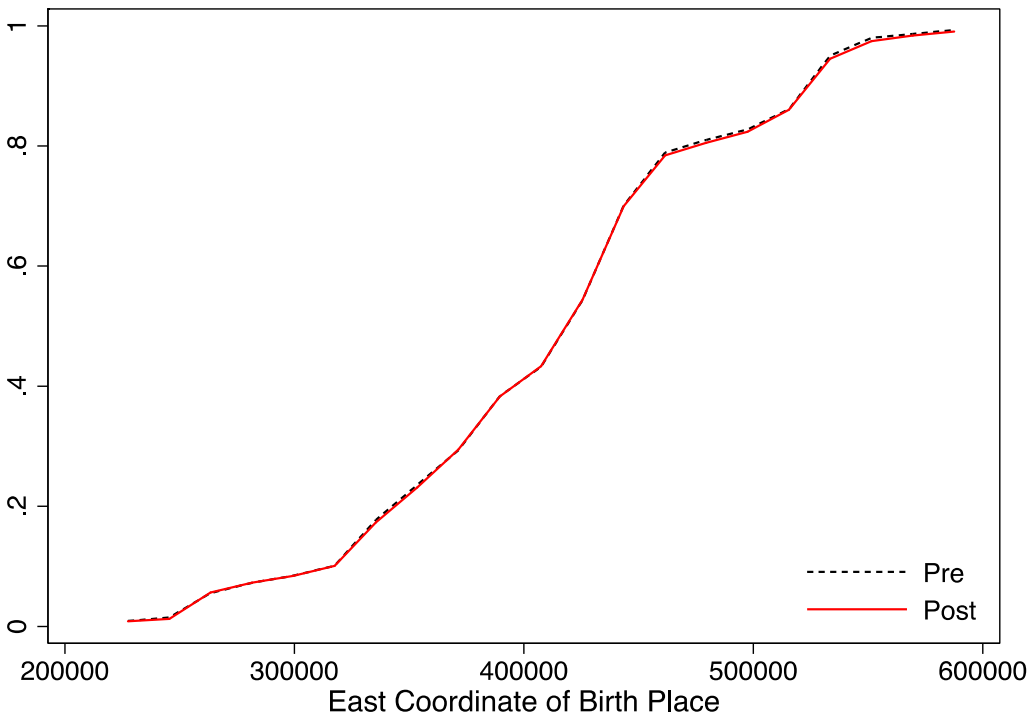
Appendix Table A1: Balance Control Test

	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

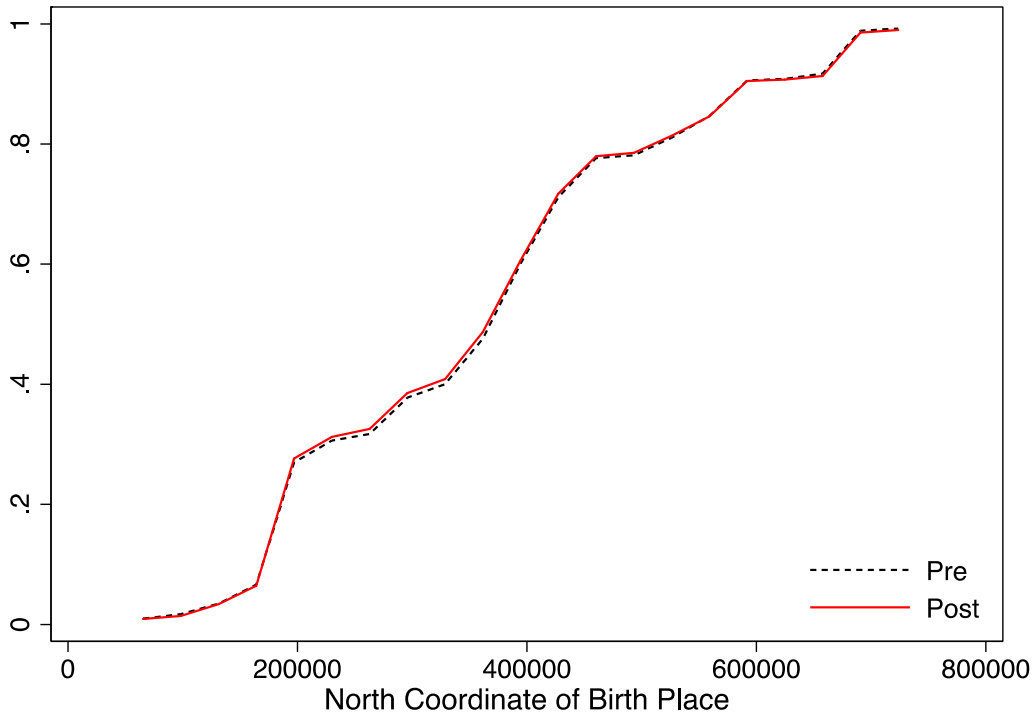
Notes: The table investigates whether predetermined characteristics are smooth are around the September 1, 1957 cutoff. It reports the coefficient on an indicator for being born on or after September 1, 1957 (i.e., "Post") from regressions where the dependent variables is listed in the column. The regressions 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



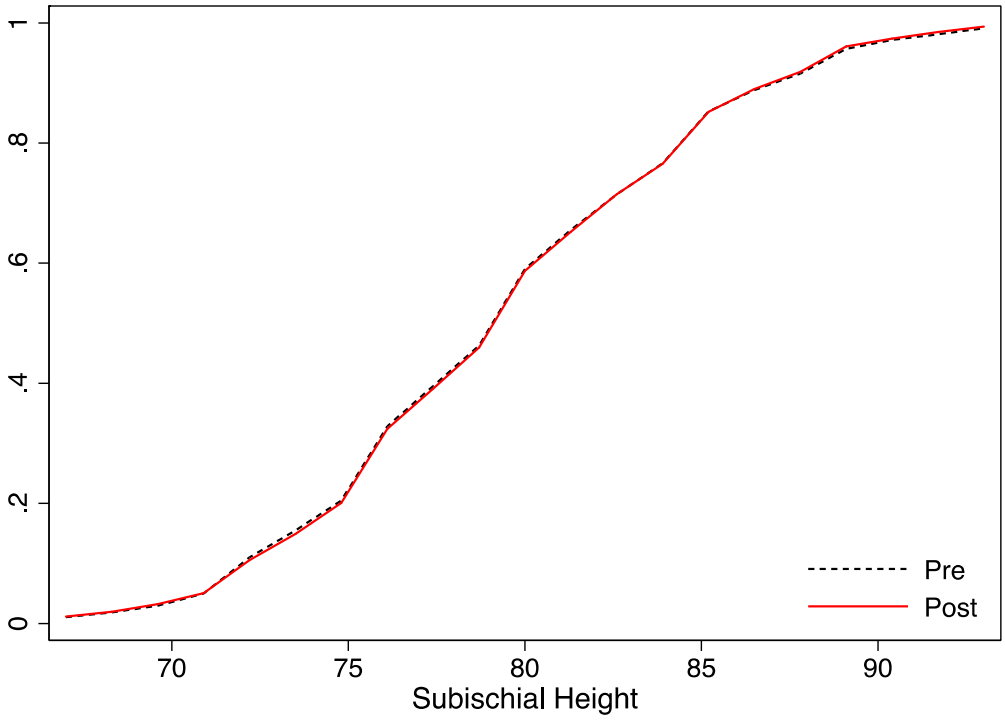
Notes: The figure shows the pre- and post-reform CDFs of east coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 266{,}883$.

Appendix Figure A17: North Coordinate of Birth Place



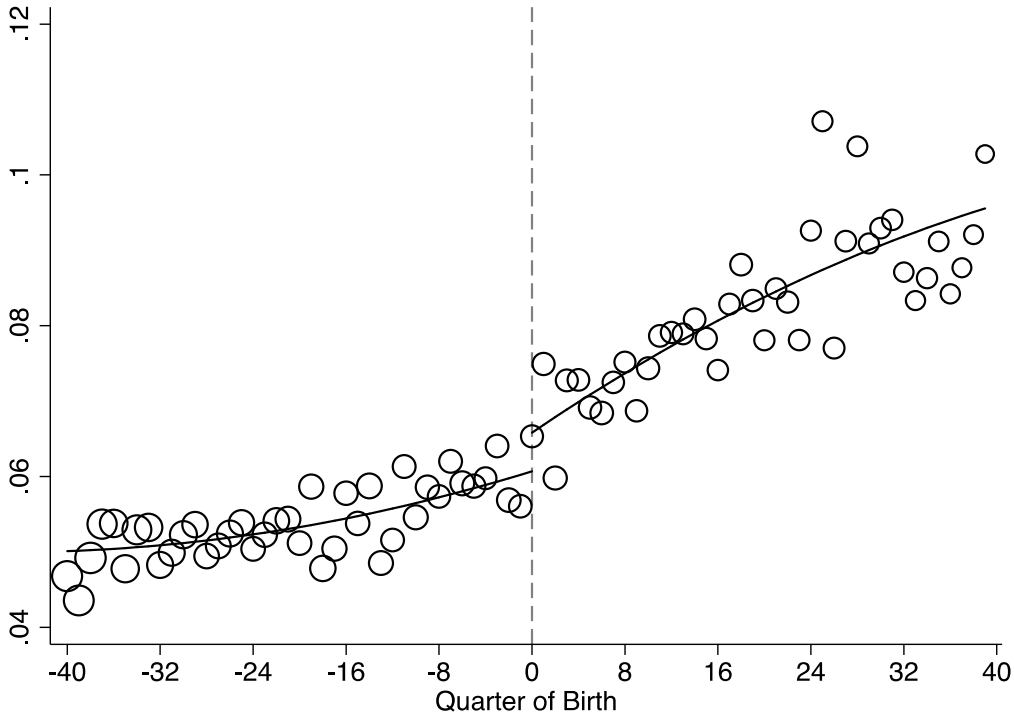
Notes: The figure shows the pre- and post-reform CDFs of north coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 266{,}883$.

Appendix Figure A18: Subischial Height



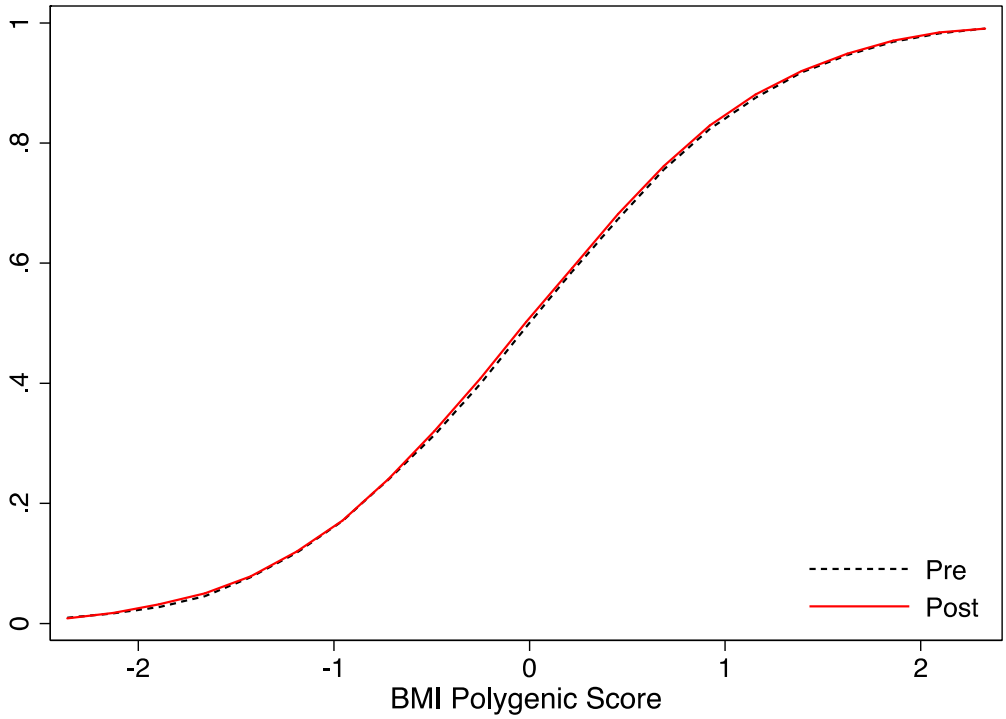
Notes:
The figure shows the pre- and post-reform CDFs of subischial height. Subischial height is the difference between standing height and sitting height. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. N = 271,173.

Appendix Figure A19: Fraction Missing Genetic Data



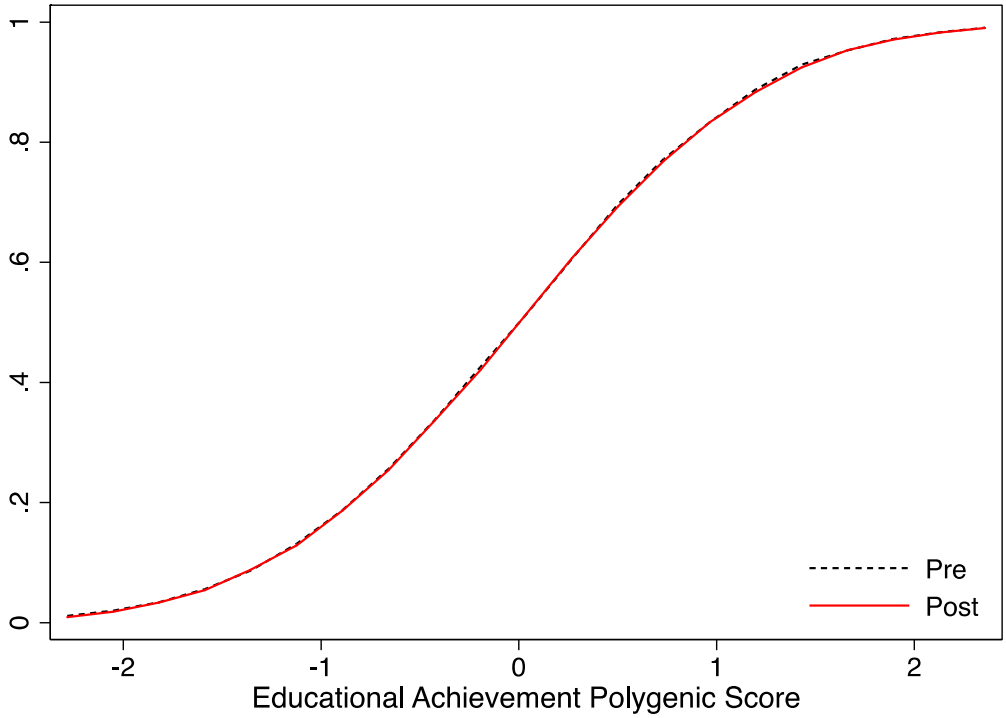
Notes: The figure shows the fraction of study participants with genetic data available by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born 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 0.0044 with a standard error of 0.0031 (p-value of 0.14). The mean among those born in the 12 months before the cutoff is 0.0591. N = 271,234.

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. $N = 253{,}715$.

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. $N = 253{,}715$.

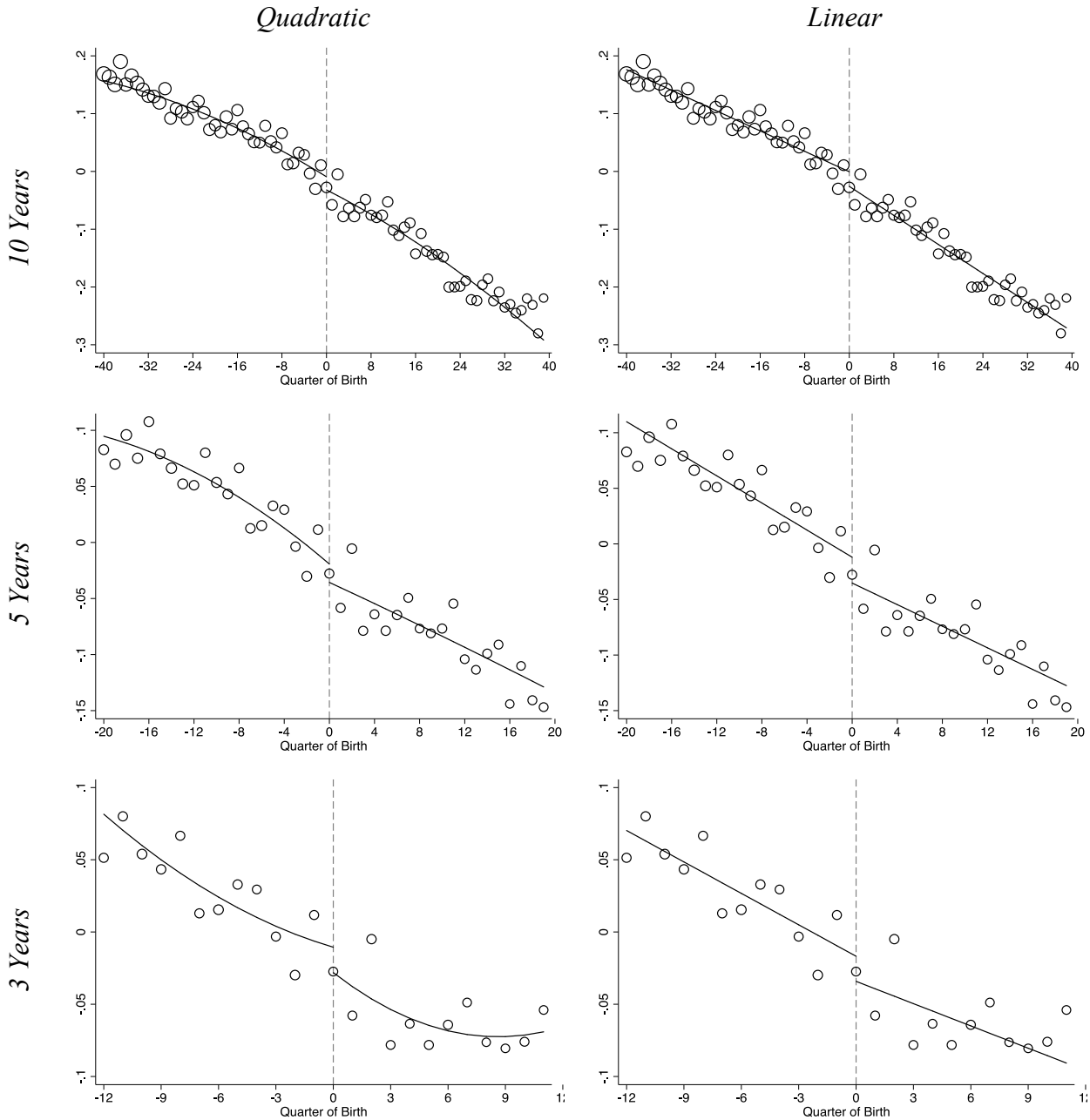
Appendix Table A2: Distributional Test

Coordinates of Birth Place			Polygenic Scores	
East	*North*	*Subischial Height*	*BMI*	*Educational Achievement*
0.65	0.21	0.63	0.32	0.92

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 253,715 (polygenic scores for BMI and educational achievement)

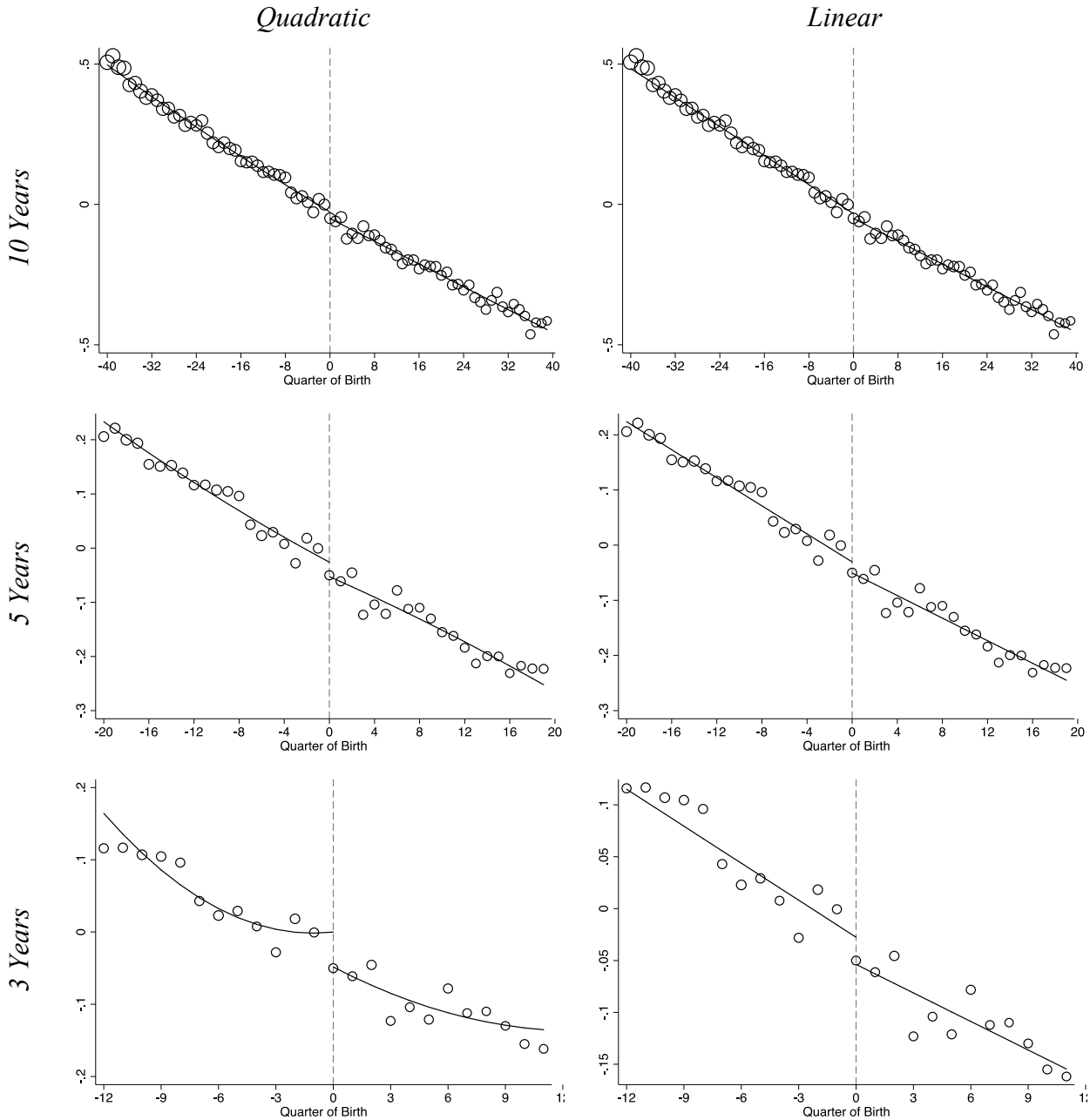
Appendix B

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



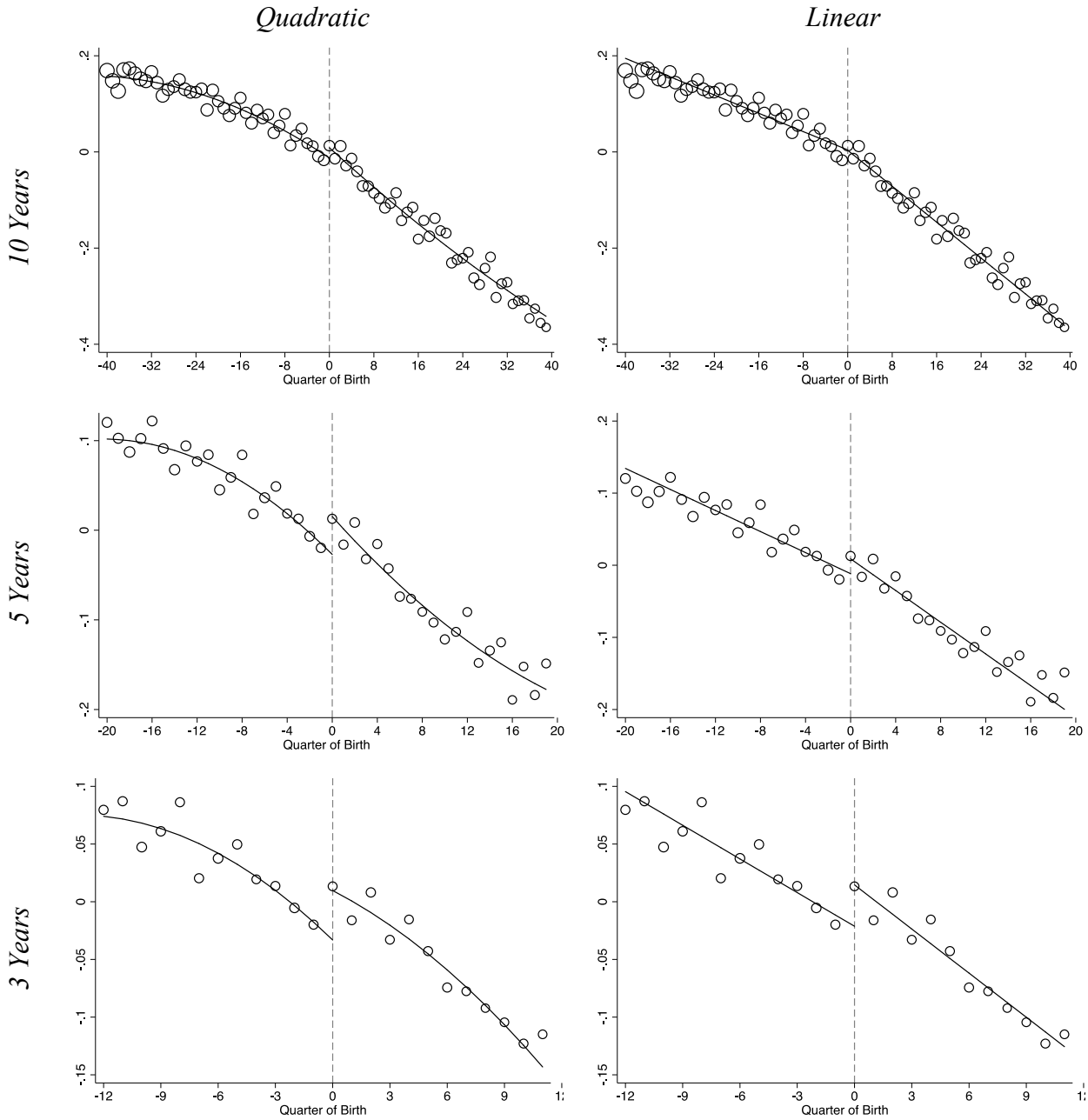
Notes: The figure assesses the sensititivity of the results for body size index to the choice of bandwidth and to the use of linear trends. It shows the average of body size index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses 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



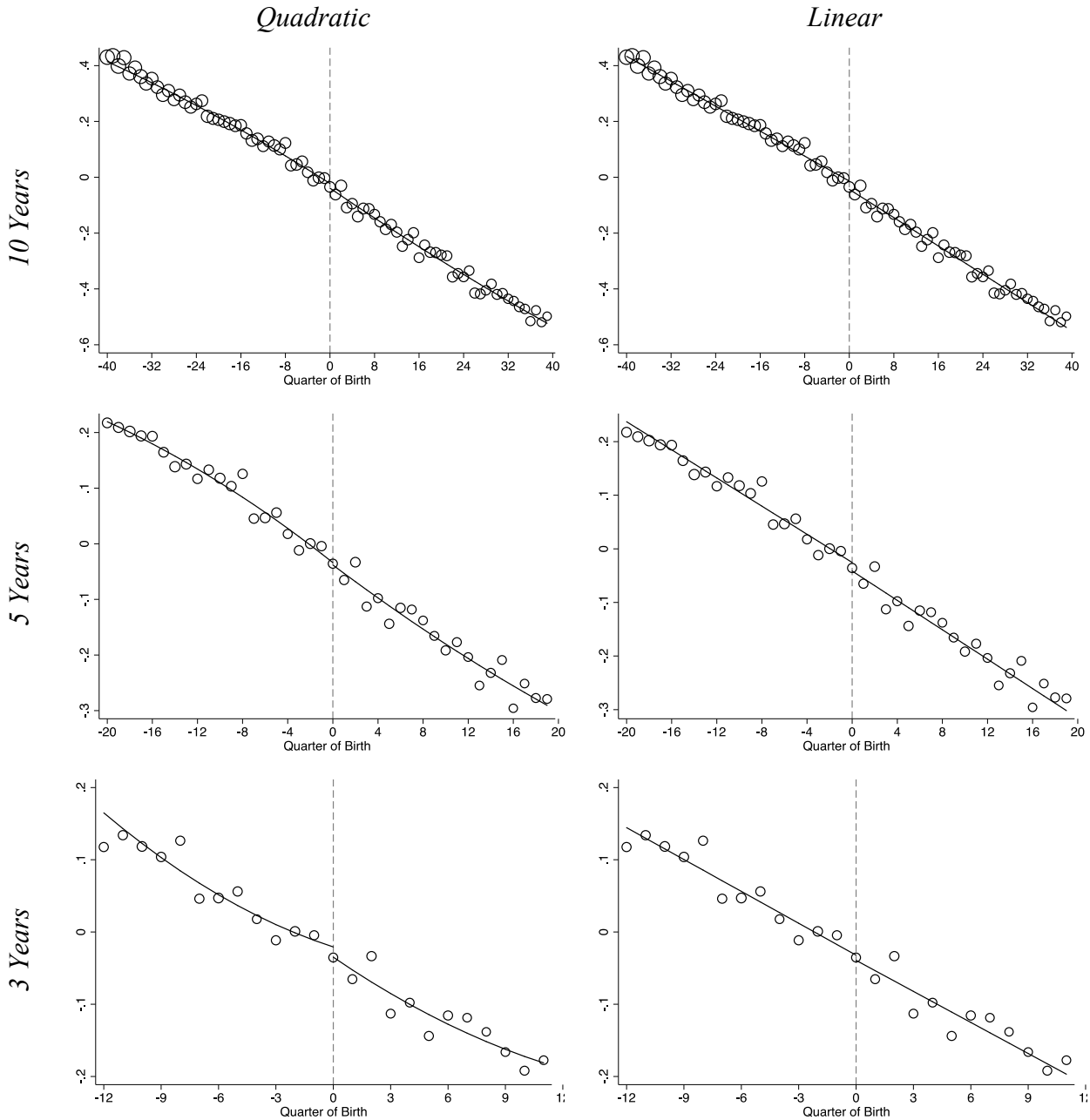
Notes: The figure assesses the sensititivity of the results for lung function index to the choice of bandwidth and to the use of linear trends. It shows the average of lung function index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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



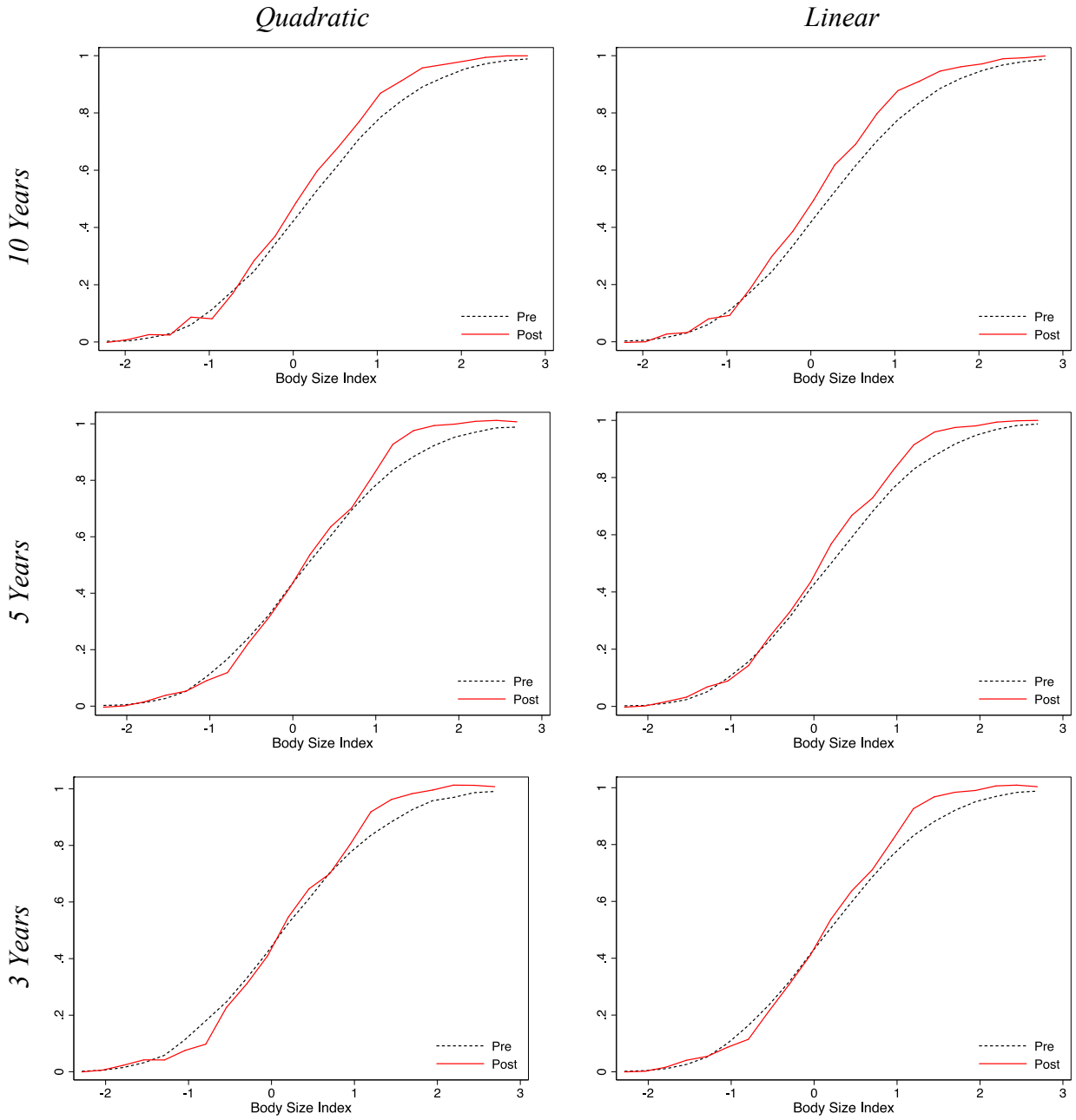
Notes: The figure assesses the sensititivity of the results for blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the average of blood pressure index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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



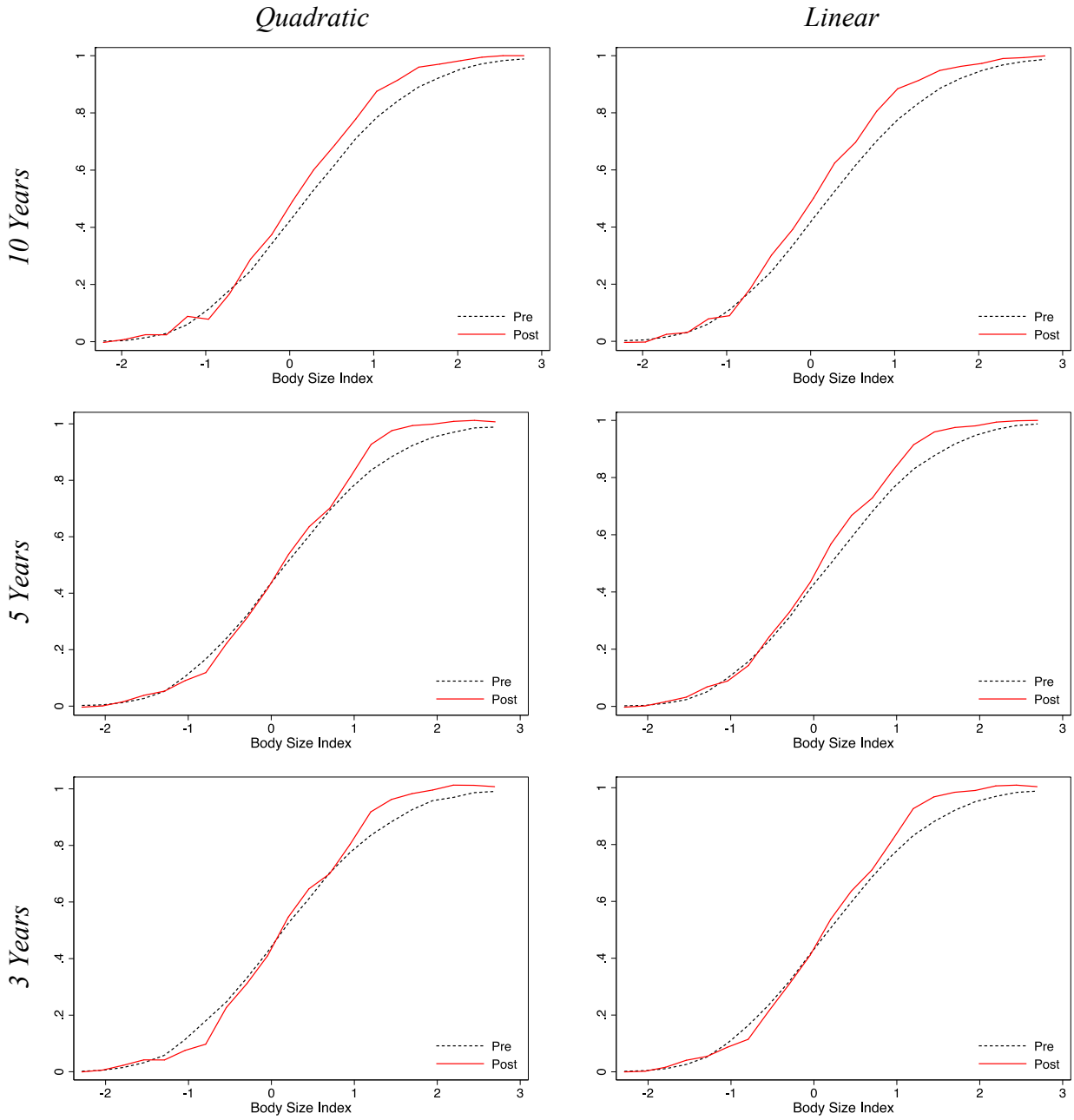
Notes: The figure assesses the sensititivity of the results for the summary index to the choice of bandwidth and to the use of linear trends. It shows the average of the summary index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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)



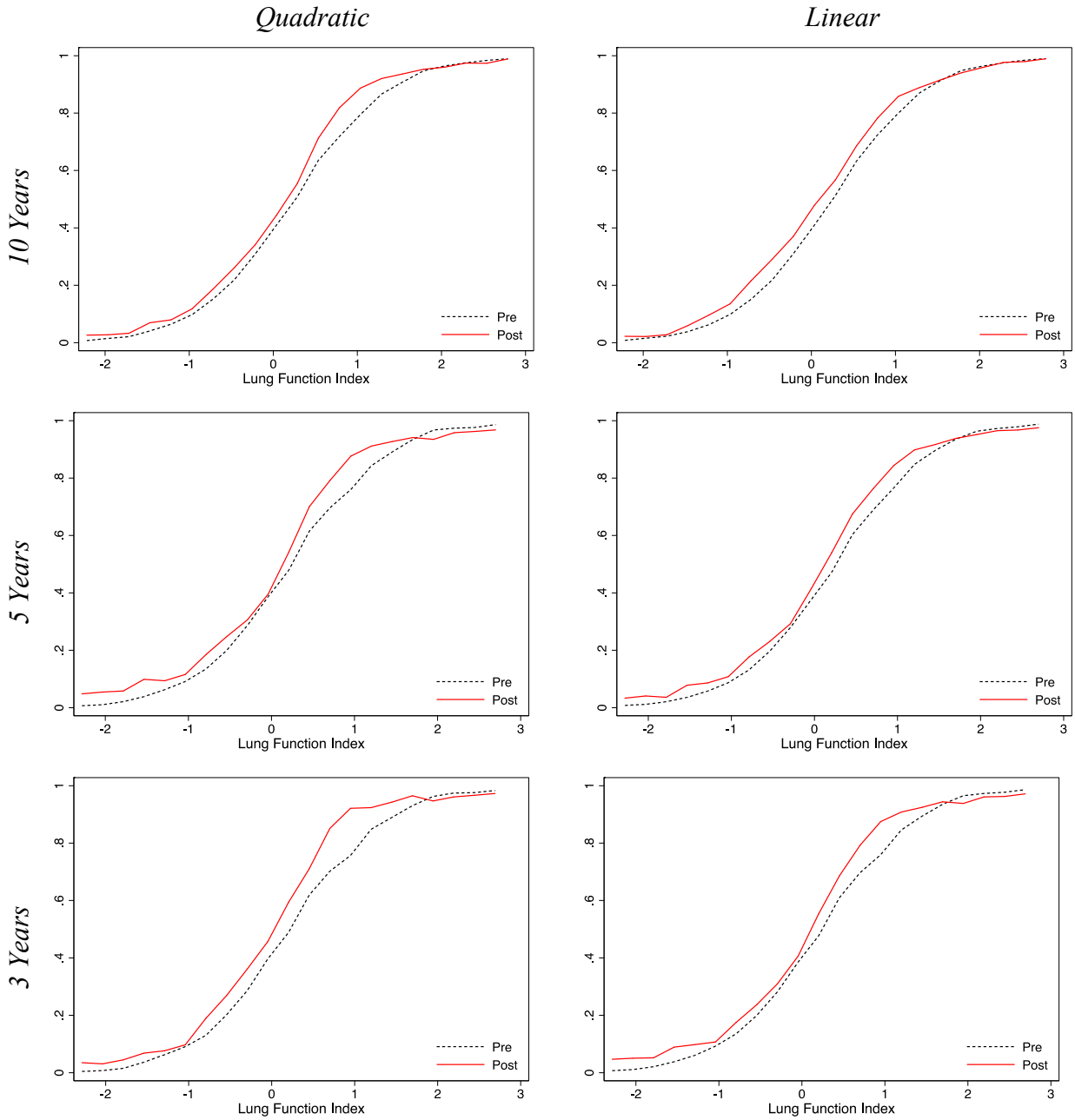
Notes: The figure assesses the sensititivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The 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. No controls.

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



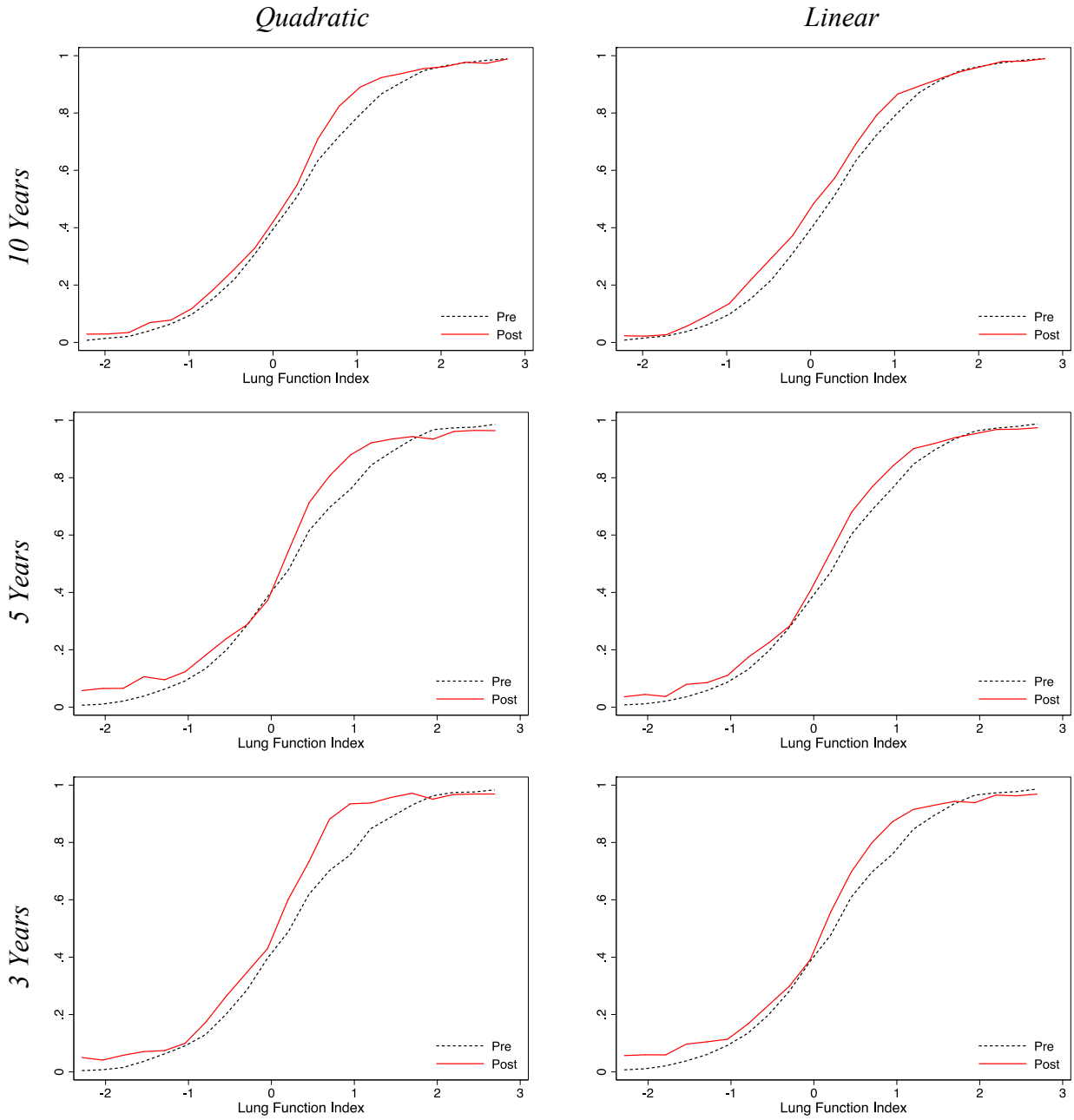
Notes: The figure assesses the sensititivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The 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)



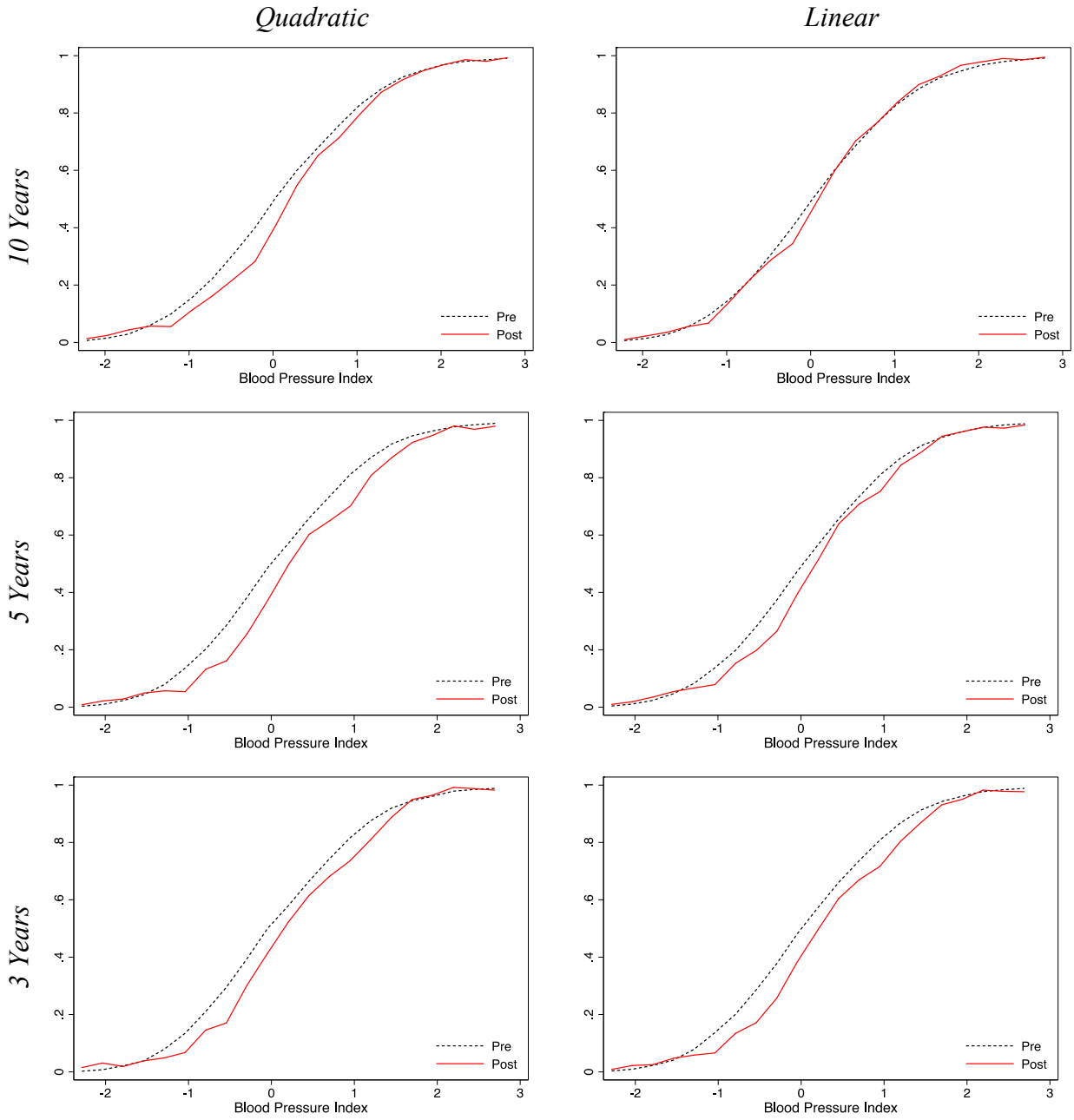
Notes: The figure assesses the sensititivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the lung function index. The left-hand side column uses quadratic trends in date of 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)



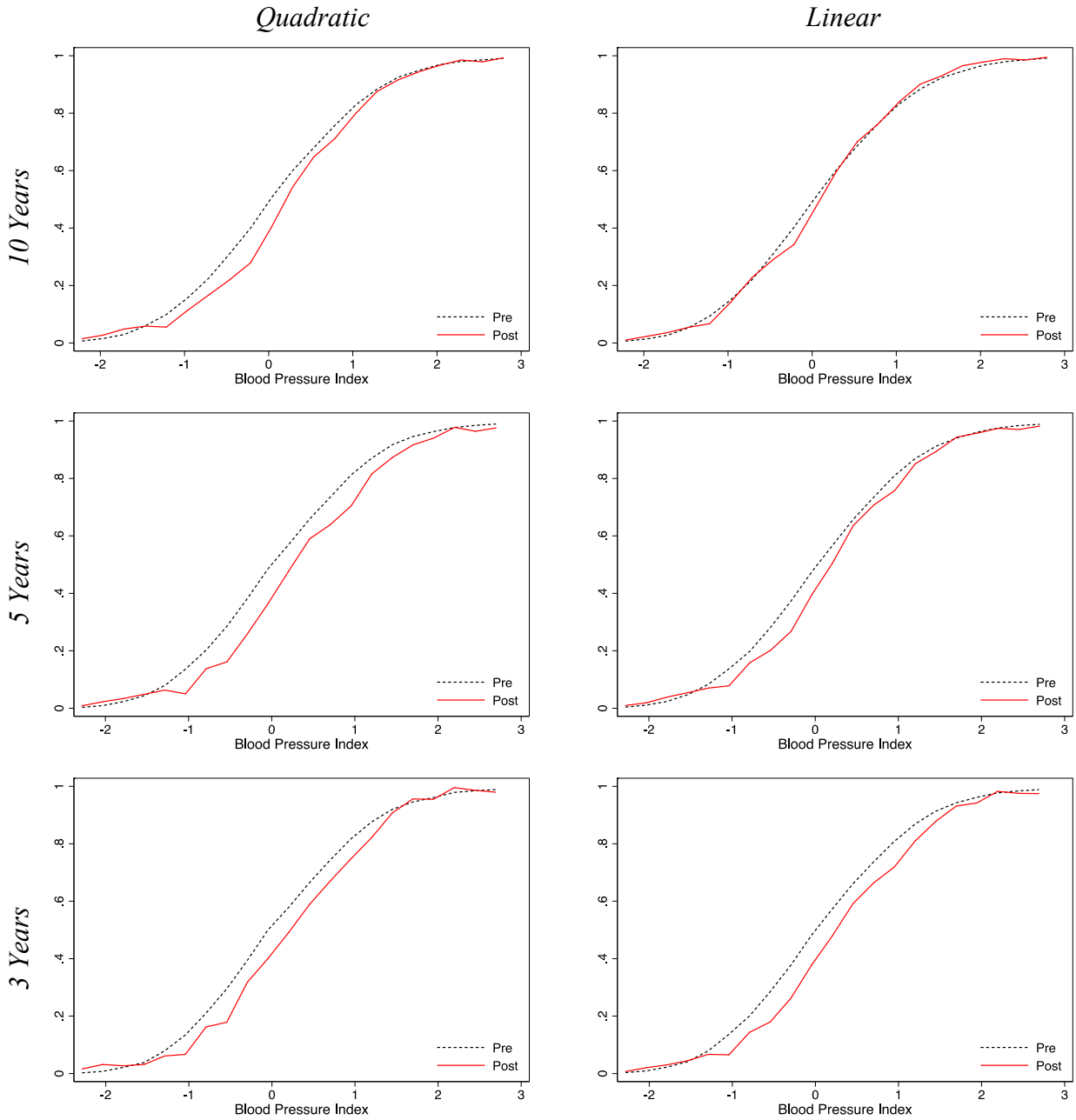
Notes: The figure assesses the sensititivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the lung function index. The left-hand side column uses quadratic trends in date of 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 Table B1: P-values of Distributional Tests for Body Size

	3 Years		5 Years		10 Years	
	Linear	*Quad.*	*Linear*	*Quad.*	*Linear*	*Quad.*
Full Distribution						
No Controls	0.2164	0.5310	0.0668	0.3070	0.0060	0.0896
With Controls	0.1656	0.3780	0.0674	0.2598	0.0050	0.0932
Bottom Half						
No Controls	0.9660	0.8154	0.9602	0.9730	0.4158	0.9526
With Controls	0.8186	0.4496	0.9716	0.8514	0.3812	0.9396
Top Half						
No Controls	0.0554	0.3446	0.0094	0.1058	0.0002	0.0126
With Controls	0.0450	0.3266	0.0094	0.0950	0.0002	0.0138

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 of the body size index.

Appendix Table B2: P-values of Distributional Tests for Lung Function

	3 Years		5 Years		10 Years	
	Linear	*Quad.*	*Linear*	*Quad.*	*Linear*	*Quad.*
Full Distribution						
No Controls	0.1354	0.1234	0.2012	0.1964	0.0768	0.1712
With Controls	0.1578	0.1572	0.2620	0.2340	0.0706	0.2352
Bottom Half						
No Controls	0.3672	0.4248	0.4022	0.3768	0.0552	0.5962
With Controls	0.3744	0.6082	0.5066	0.4406	0.0626	0.7634
Top Half						
No Controls	0.0626	0.0438	0.1102	0.1168	0.1090	0.0618
With Controls	0.0778	0.0524	0.1402	0.1396	0.0754	0.0870

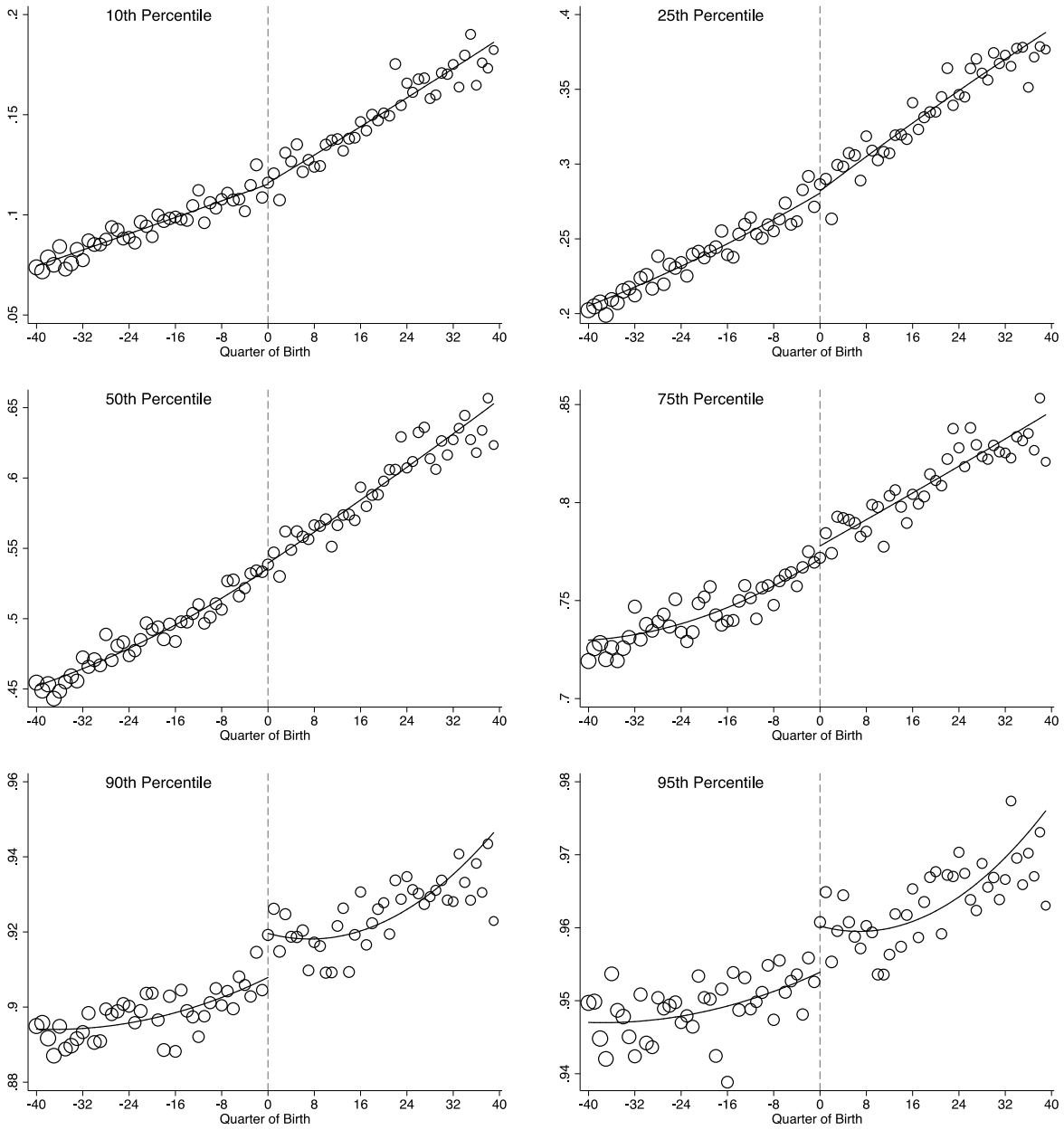
Notes: The table shows the p-values of tests of the equality of the full distribution, the bottom and top halves of the pre- and post-reform CDFs of the lung function index.

Appendix Table B3: P-values of Distributional Tests for Blood Pressure

	3 Years		5 Years		10 Years	
	Linear	*Quad.*	*Linear*	*Quad.*	*Linear*	*Quad.*
Full Distribution						
No Controls	0.0208	0.1480	0.0414	0.0262	0.5432	0.0362
With Controls	0.0306	0.2006	0.0552	0.0358	0.5856	0.0532
Bottom Half						
No Controls	0.0100	0.0836	0.0112	0.0126	0.4152	0.0102
With Controls	0.0226	0.1806	0.0196	0.0266	0.4658	0.0172
Top Half						
No Controls	0.0420	0.2772	0.1984	0.0500	0.6924	0.1502
With Controls	0.0380	0.2264	0.2008	0.0514	0.7240	0.1684

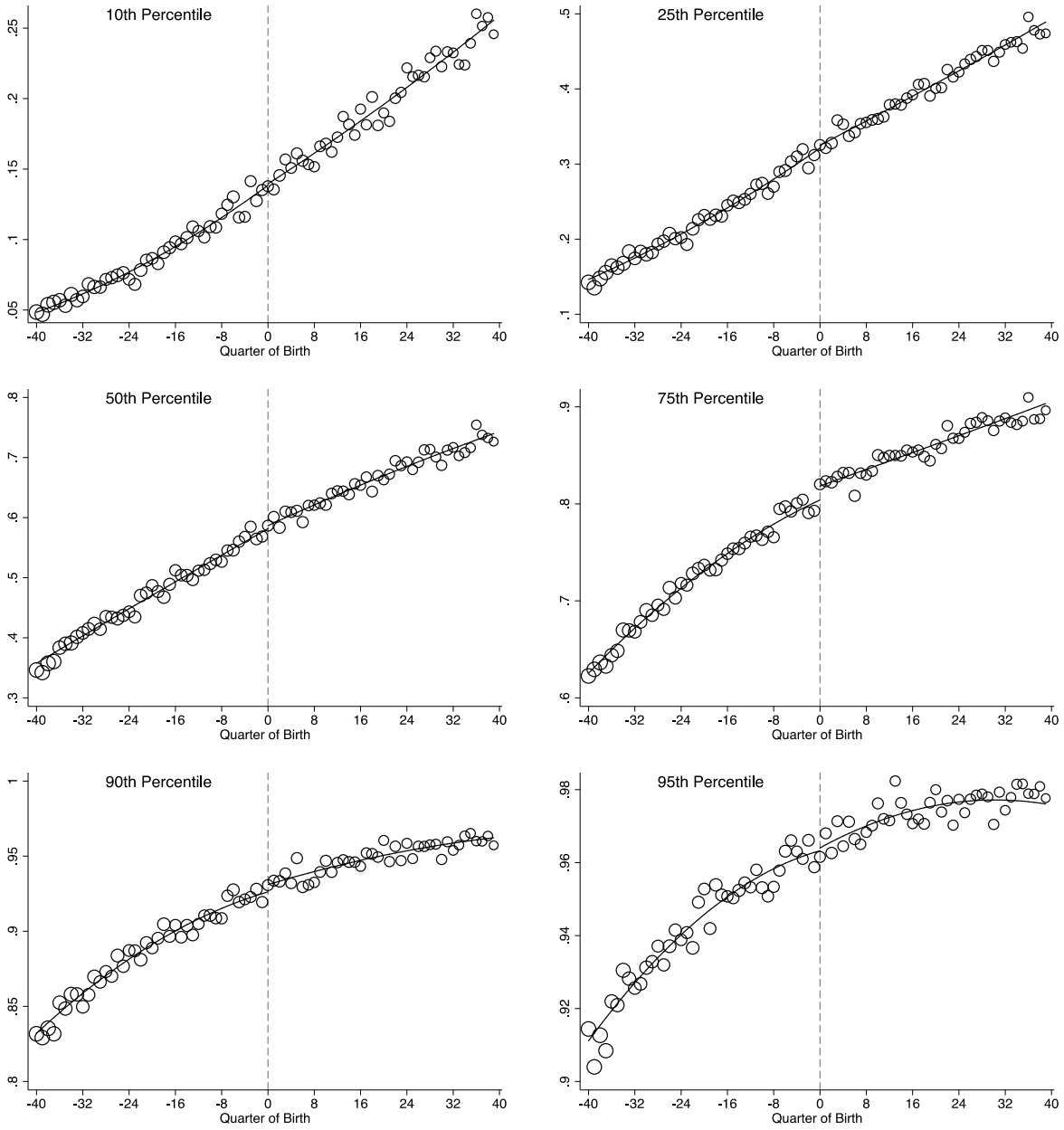
Notes: The table shows the p-values of tests of the equality of the full distribution, the bottom and top halves of the pre- and post-reform CDFs of the blood pressure index.

Appendix Figure B11: Effect on Percentiles of Distribution of Body Size Index



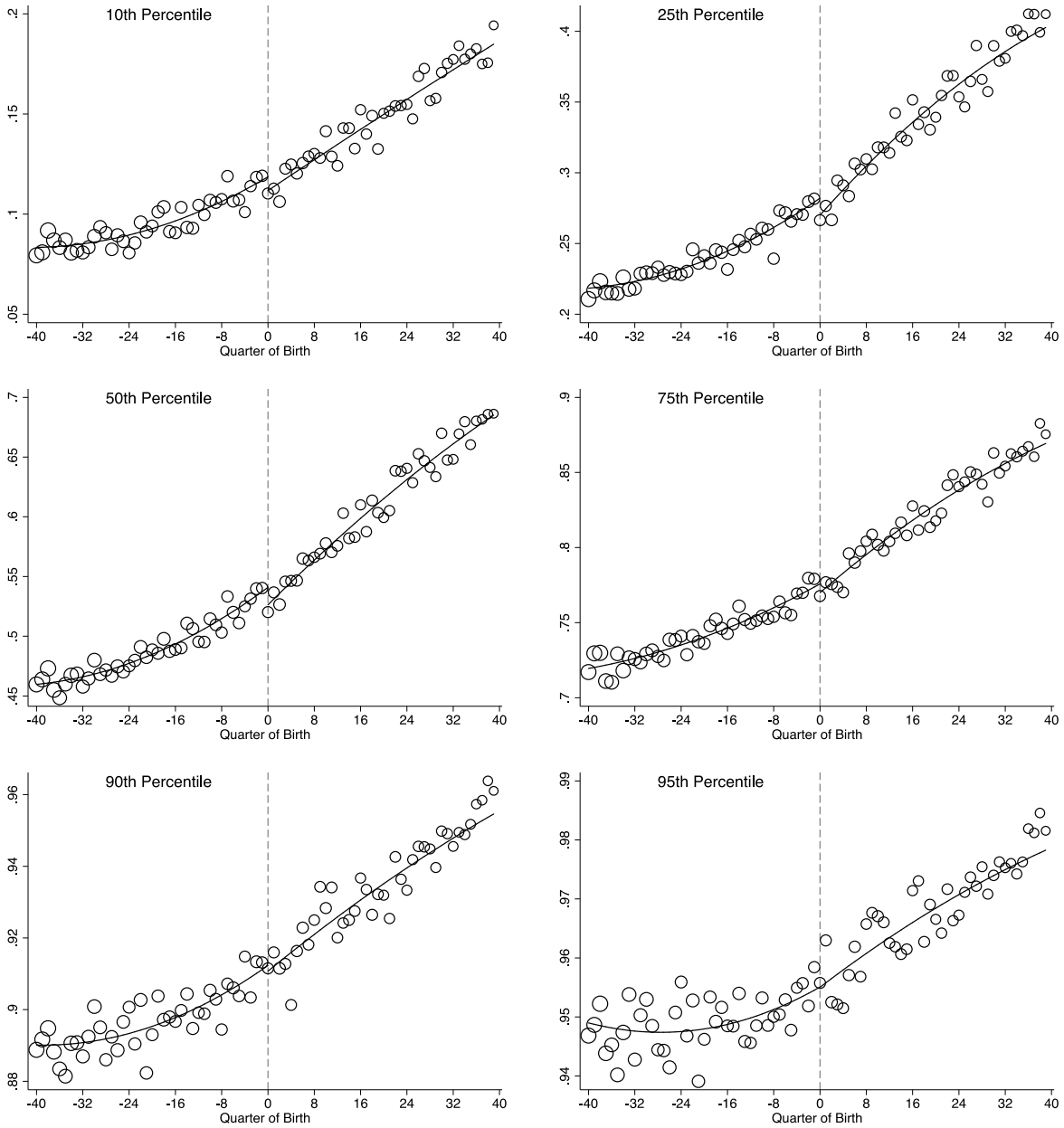
Notes: The figure shed lights on the results shown in Figure 4 of the paper. It shows the fraction of study participants with a body size index below the 10th, 25th, 50th, 75th, 90th, and the 95th percentile (of the distribution of those born between September 1, 1956 and August 31, 1957) by quarter of birth. 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. $N = 266{,}525$.

Appendix Figure B12: Effect on Percentiles of Distribution of Lung Function Index



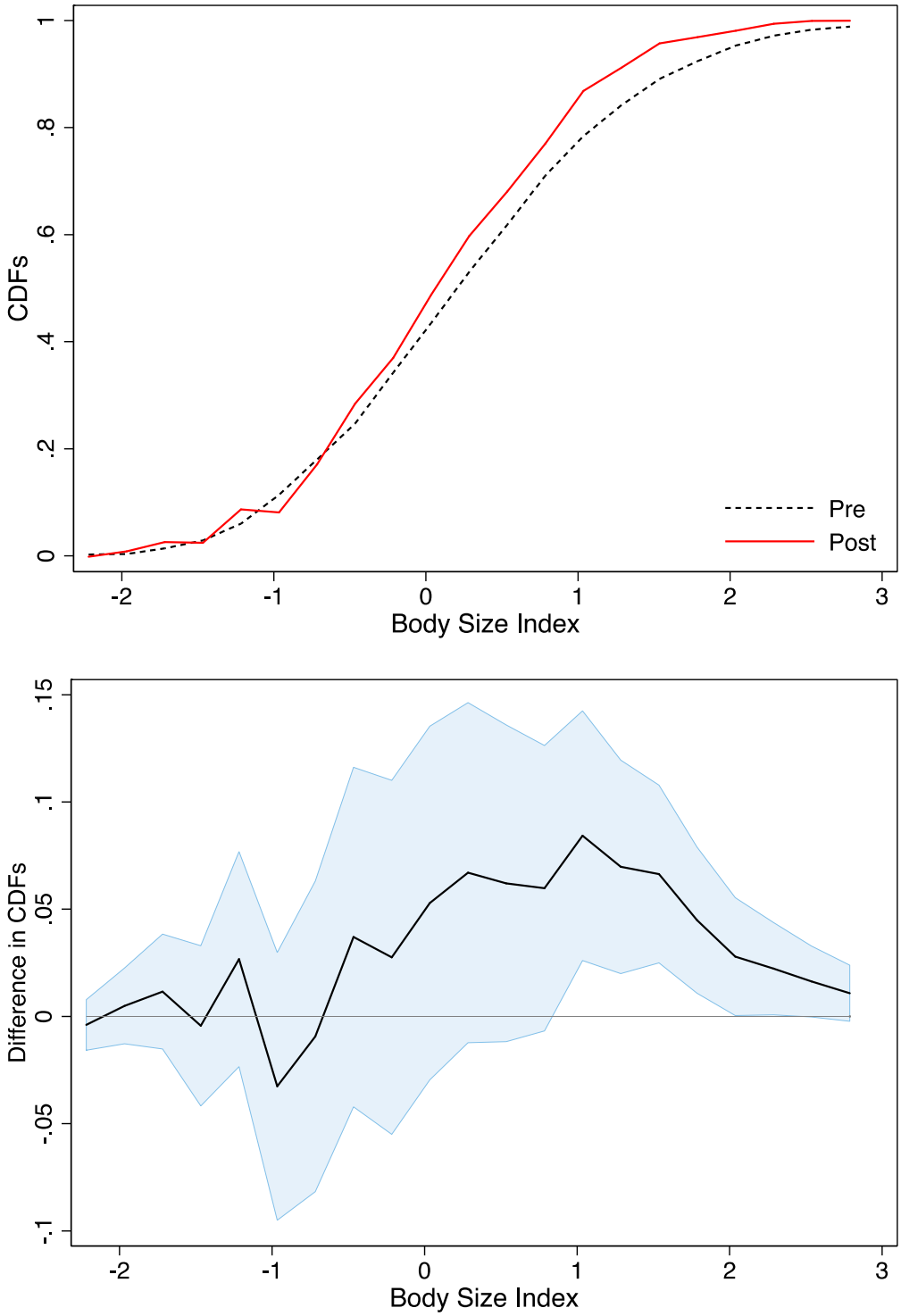
Notes: The figure shed lights on the results shown in Figure 5 of the paper. It shows the fraction of study participants with a body size index below the 10th, 25th, 50th, 75th, 90th, and the 95th percentile (of the distribution of those born between September 1, 1956 and August 31, 1957) by quarter of birth. 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 = 215,366.

Appendix Figure B13: Effect on Percentiles of Distribution of Blood Pressure Index



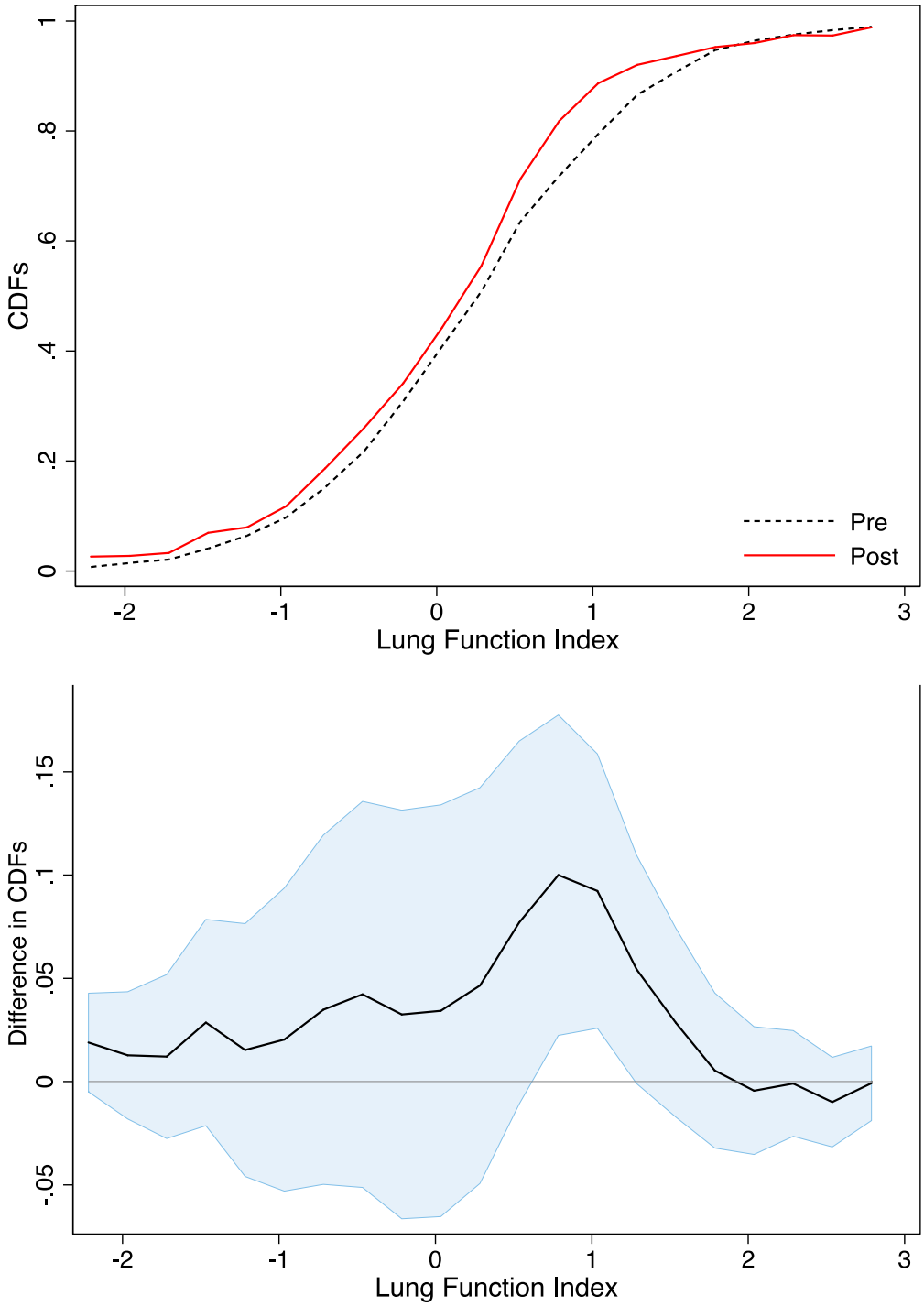
Notes: The figure shed lights on the results shown in Figure 6 of the paper. It shows the fraction of study participants with a body size index below the 10th, 25th, 50th, 75th, 90th, and the 95th percentile (of the distribution of those born between September 1, 1956 and August 31, 1957) by quarter of birth. 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 = 270,647.

Appendix Figure B14: Distributional Effects on Body Size Index



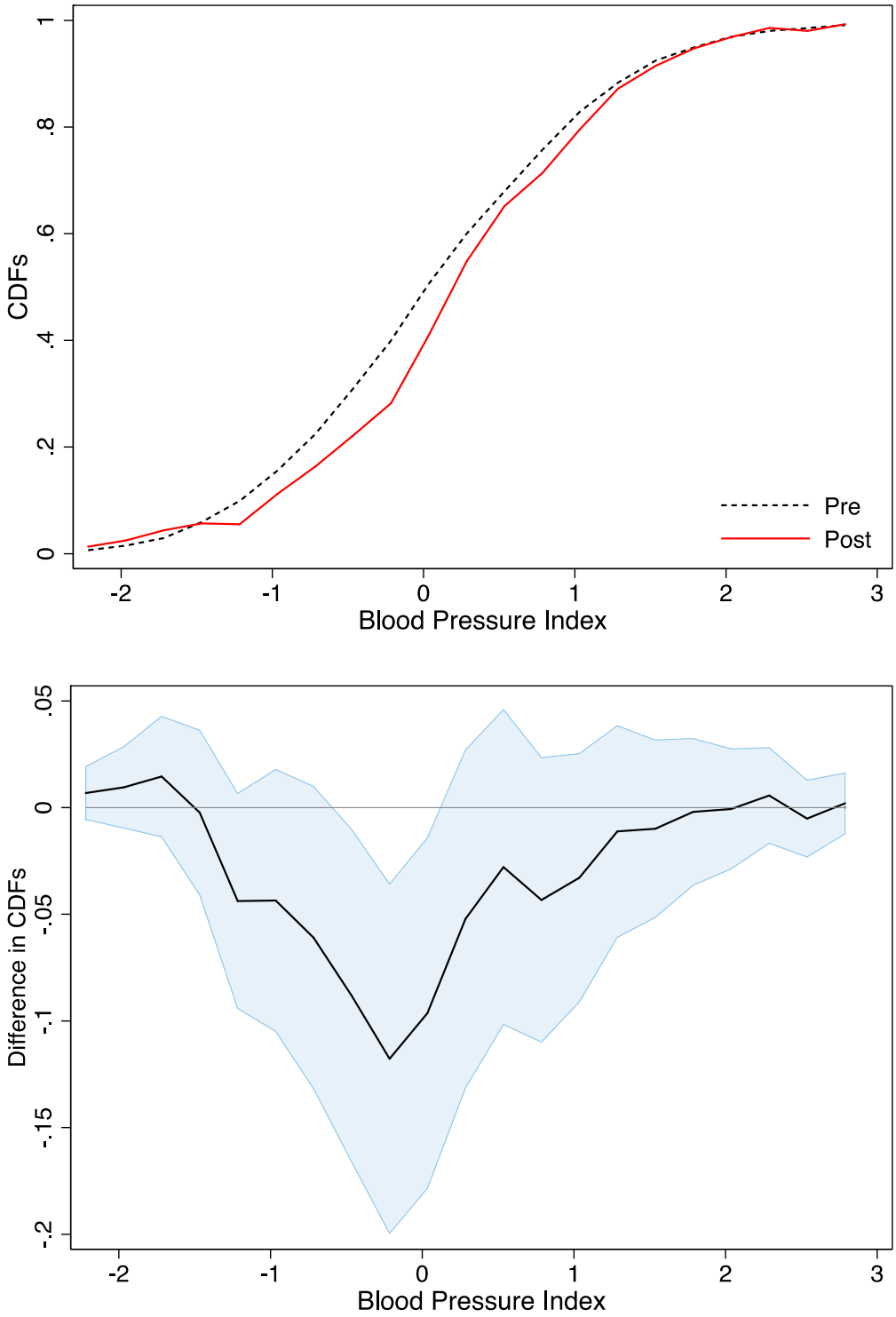
Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 4 in the paper, showing the pre- and post-reform CDF of body size index for compliers. The black solid line in the bottom figure shows the difference between 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



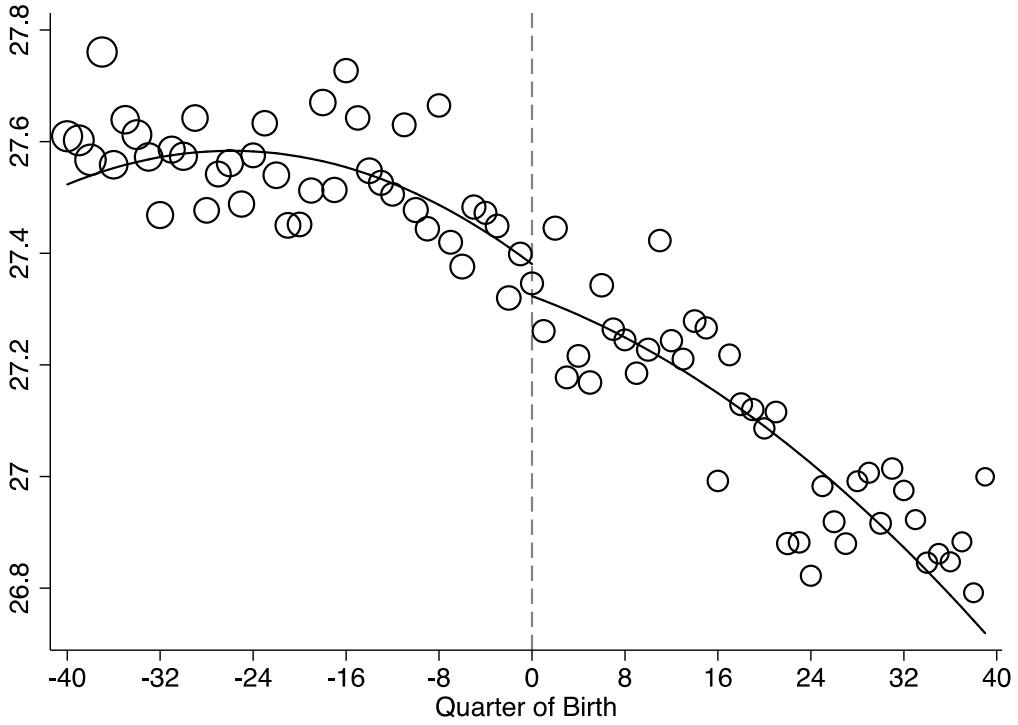
Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 5 in the paper, showing the pre- and post-reform CDF of lung function index for compliers. The black solid line in the bottom figure shows the difference 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



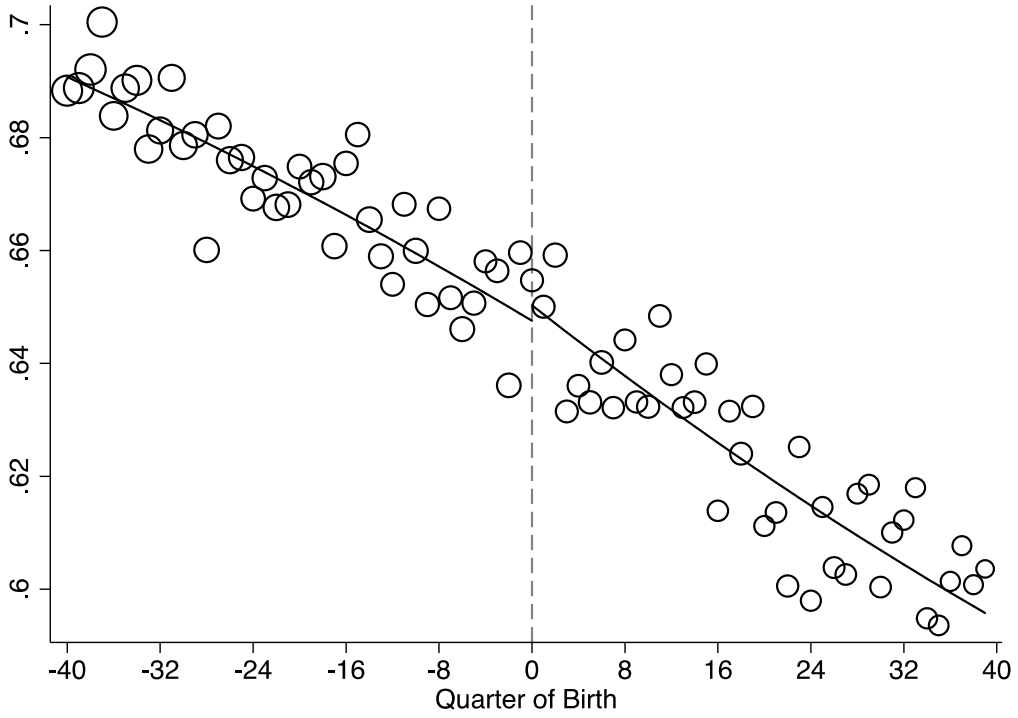
Notes: The figure shows the difference between the pre- and post-reform CDFs for compliers and 95% confidence bands. The top figure reproduces Figure 6 in the paper, showing the pre- and post-reform CDF of blood pressure index for compliers. The black solid line in the bottom figure shows the difference 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 Figure B16: Body Mass Index



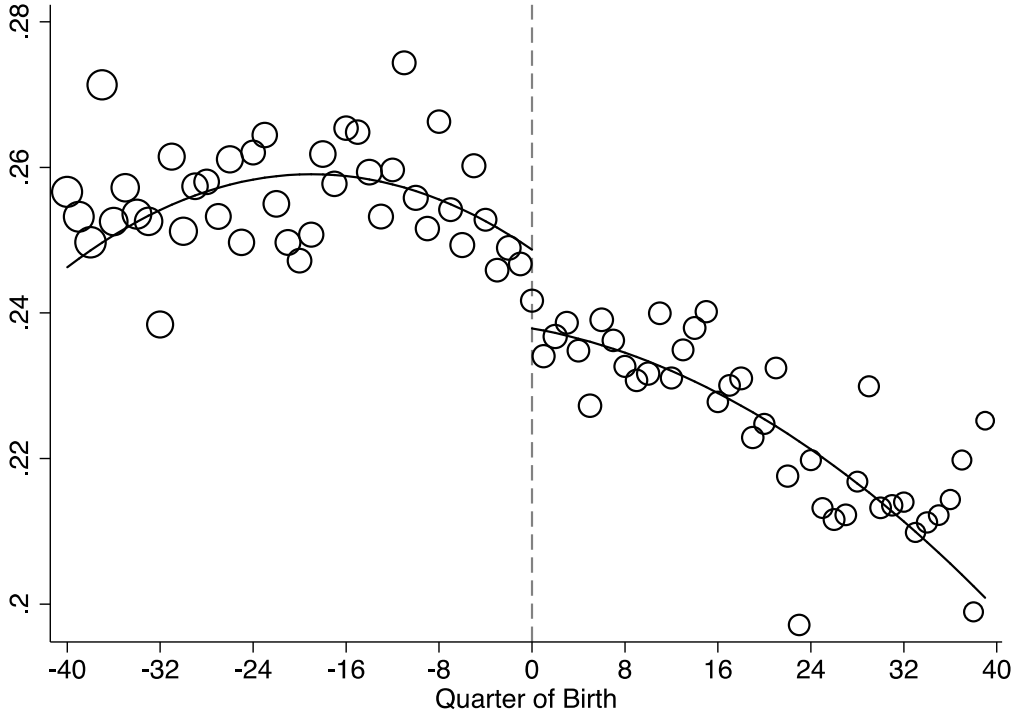
Notes: The figure shows average BMI by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show quadratic 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 = 270,019.

Appendix Figure B17: Overweight



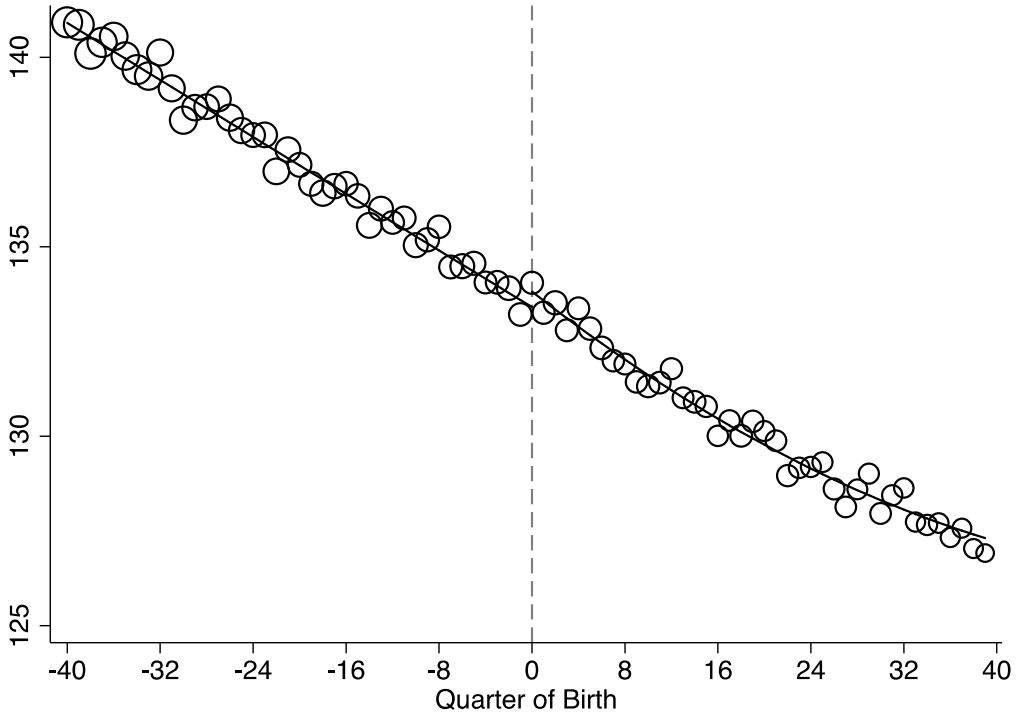
Notes: The figure shows the fraction of study participants who were overweight by quarter of birth. Overweight is defined as having a BMI greater or equal to 25. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line 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 = 270,019.

Appendix Figure B18: Obese



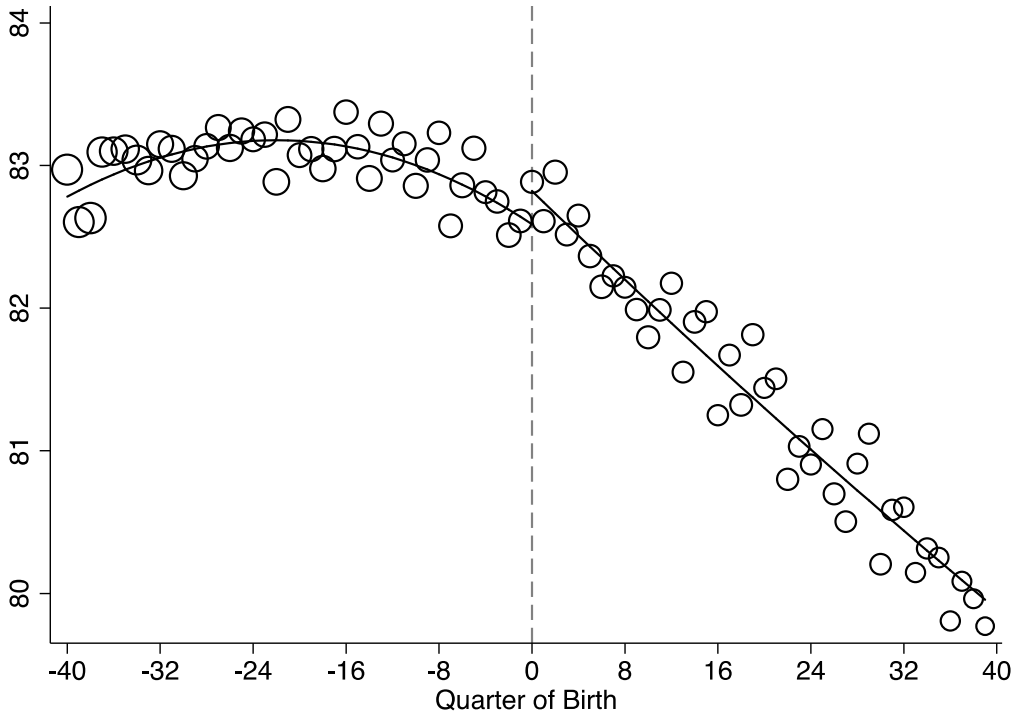
Notes: The figure shows the fraction of study participants who were obese by quarter of birth. Obesity is defined as having a BMI greater or equal to 30. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to 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. $N = 270{,}019$.

Appendix Figure B19: Blood Pressure Systolic



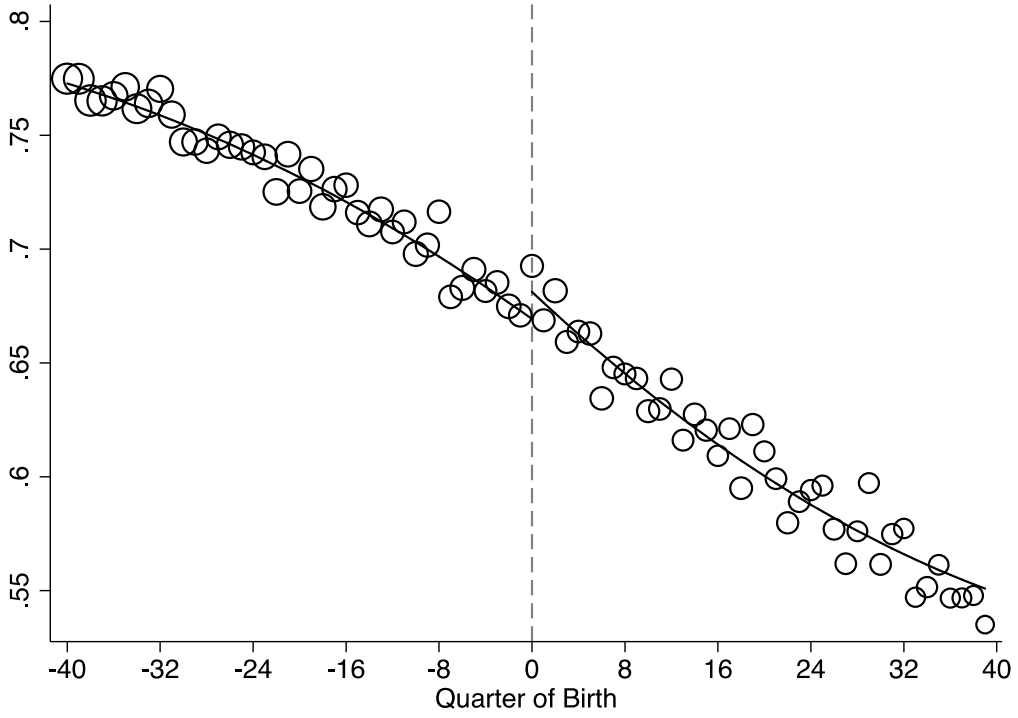
Notes: The figure shows average systolic blood pressure by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. 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. $N = 270{,}647$.

Appendix Figure B20: Diastolic Blood Pressure



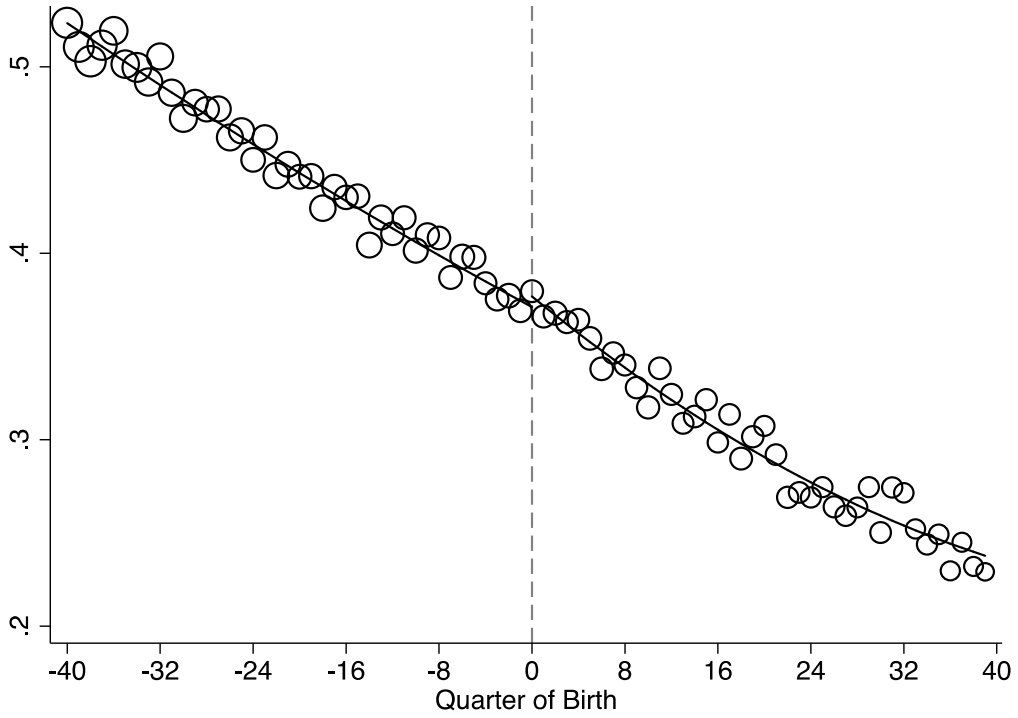
Notes: The figure shows average diastolic blood pressure by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. 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. $N = 270{,}647$.

Appendix Figure B21: Stage 1 Hypertension



Notes: The figure shows the fraction of participants with stage 1 hypertension by quarter of birth. Stage 1 hypertension is defined as having a diastolic blood pressure greater or equal to 80 or having a systolic blood pressure greater or equal to 130. 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. $N = 270{,}647$.

Appendix Figure B22: Stage 2 Hypertension



Notes: The figure shows the fraction of participants with stage 2 hypertension by quarter of birth. Stage 2 hypertension is defined as having a diastolic blood pressure greater or equal to 90 or having a systolic blood pressure greater or equal to 140. 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 = 270,647.

Appendix Table B4: BMI, Overweight, and Obesity

	BMI		*Overweight*		*Obesity*	
Reduced-form						
Post	-0.061	-0.070	0.003	0.001	-0.011	-0.012
	[0.063]	[0.063]	[0.006]	[0.006]	[0.005]**	[0.005]**
Two stages least squares						
Stayed in school until 16	-0.407	-0.504	0.020	0.009	-0.075	-0.085
	[0.420]	[0.453]	[0.040]	[0.043]	[0.036]**	[0.039]**
Controls?	No	Yes	No	Yes	No	Yes
Mean of Y	27.41	27.41	0.65	0.65	0.25	0.25
N Observations	270,019	270,019	270,019	270,019	270,019	270,019

Notes: The table shows the effects on average BMI, the fraction overweight, and the fraction obese. 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.

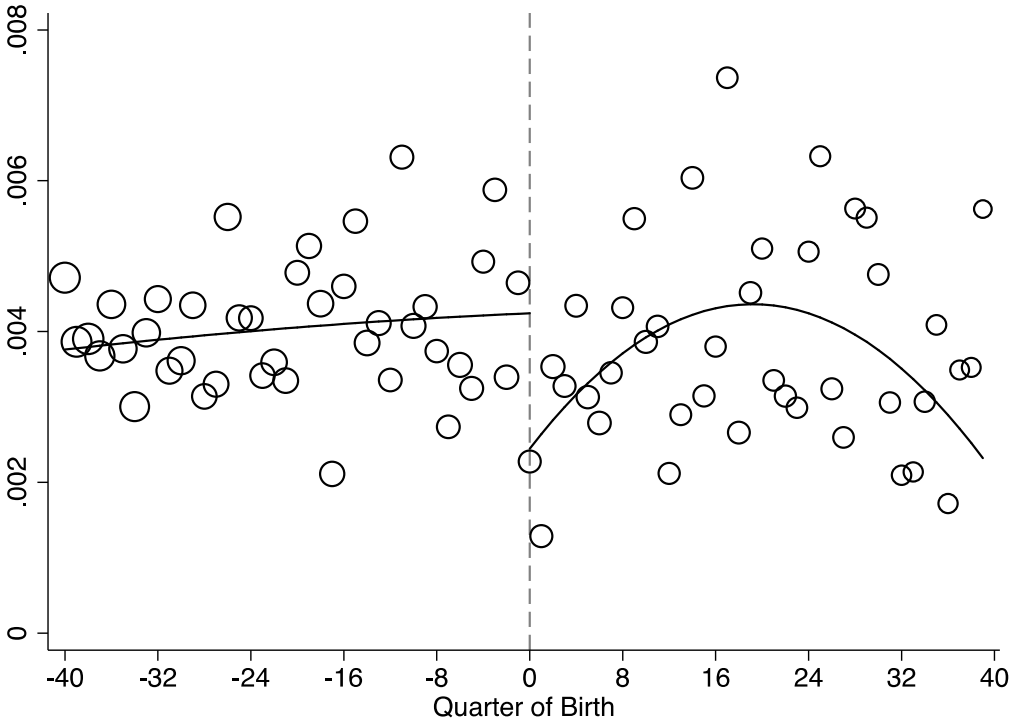
Appendix Table B5: Blood Pressure

	Systolic Blood Pressure		*Diastolic Blood Pressure*		*Stage 1 Hypertension*		*Stage 2 Hypertension*	
Reduced-form								
Post	0.426	0.311	0.243	0.208	0.012	0.010	0.006	0.005
	[0.213]**	[0.209]	[0.130]*	[0.128]	[0.006]**	[0.006]*	[0.006]	[0.006]
Two stages least squares								
Stayed in school until 16	2.836	2.234	1.619	1.492	0.082	0.071	0.041	0.037
	[1.420]**	[1.508]	[0.866]*	[0.921]	[0.039]**	[0.042]*	[0.041]	[0.043]
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Y	133.80	133.80	82.66	82.66	0.68	0.68	0.38	0.38
N Observations	270,647	270,647	270,647	270,647	270,647	270,647	270,647	270,647

Notes: The table shows the effects on average systolic blood pressure, diastolic blood pressure, stage 1 hypertension, and stage 2 hypertension. Stage 1 hypertension is defined as having a systolic blood pressure greater or equal to 130 or a diastolic blood pressure greater or equal to 80. Stage 2 hypertension is defined as having a systolic blood pressure greater or equal to 140 or a diastolic blood pressure greater or equal to 90. 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.

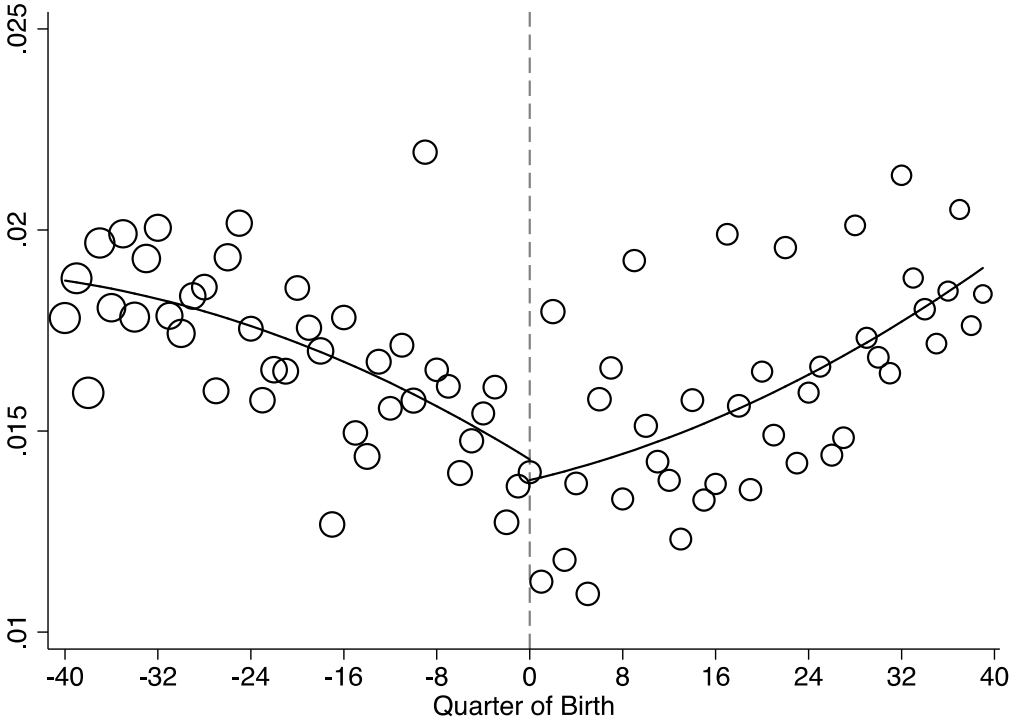
Appendix C

Appendix Figure C1: Missing Body Mass Index



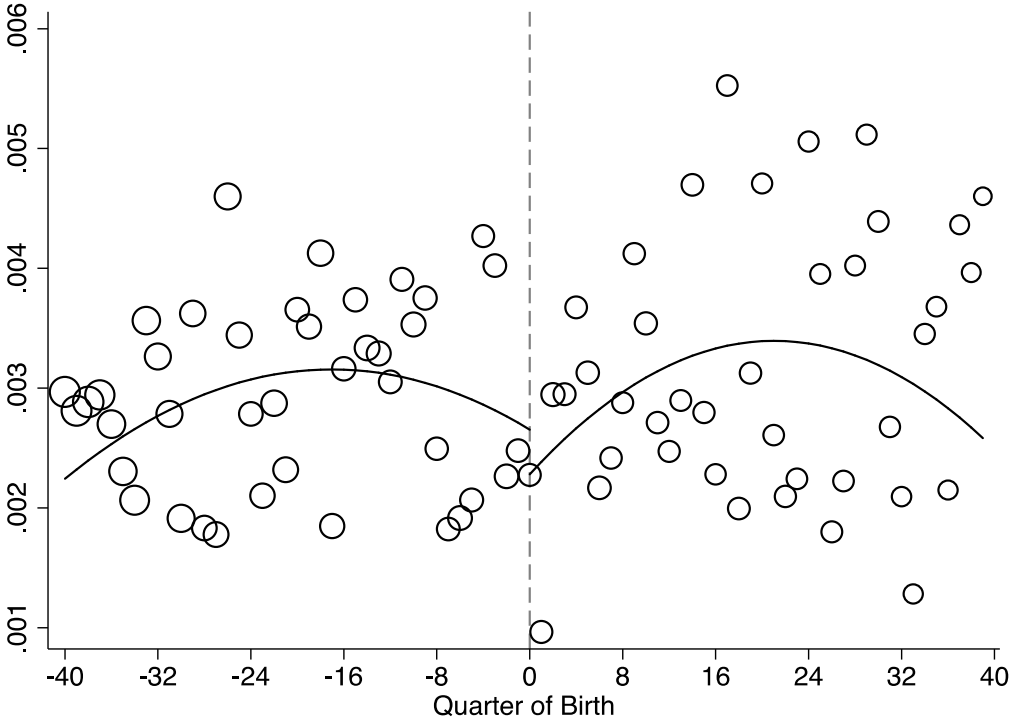
Notes: The figure shows the fraction of study participants for whom data on BMI was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts 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



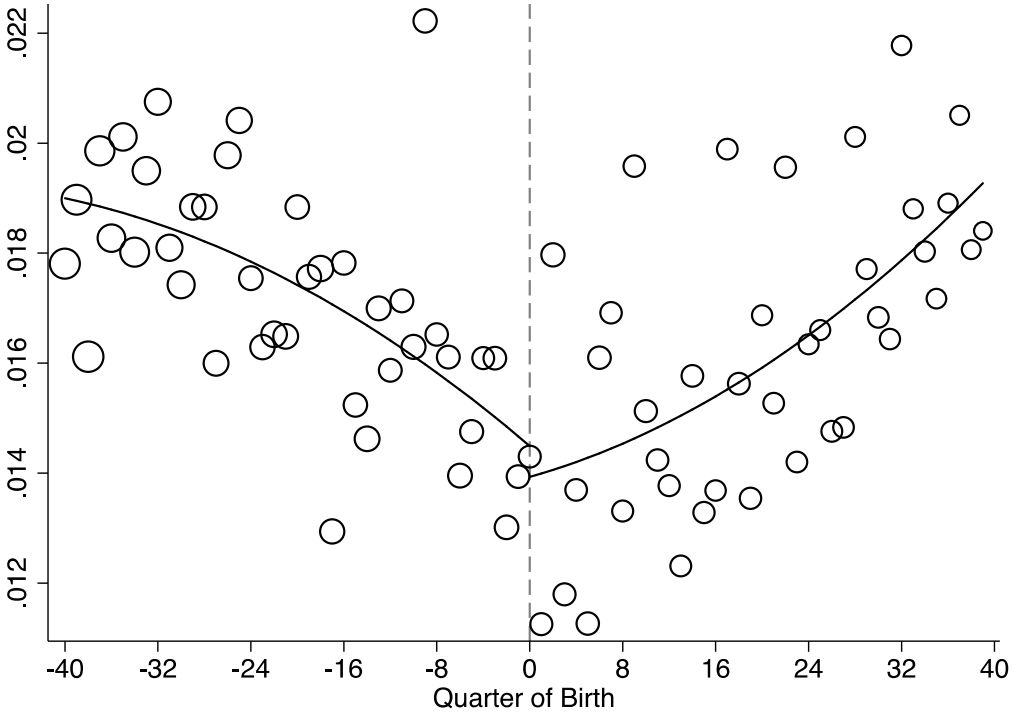
Notes: The figure shows the fraction of study participants for whom data on body fat percentage was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 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



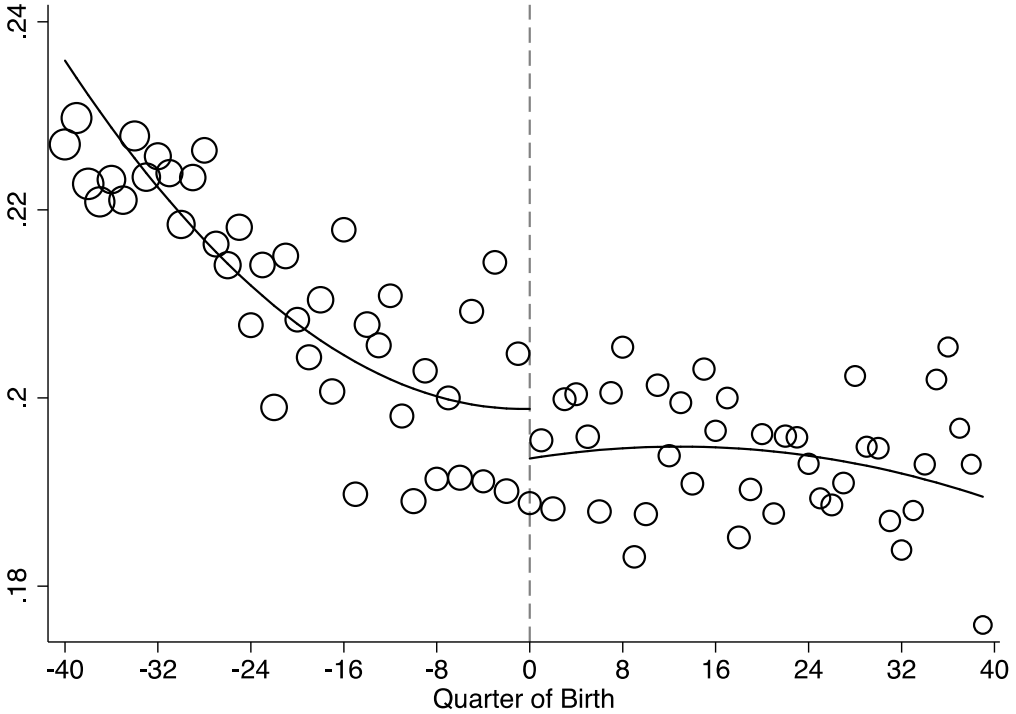
Notes: The figure shows the fraction of study participants for whom data on waist-hip ratio was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while 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



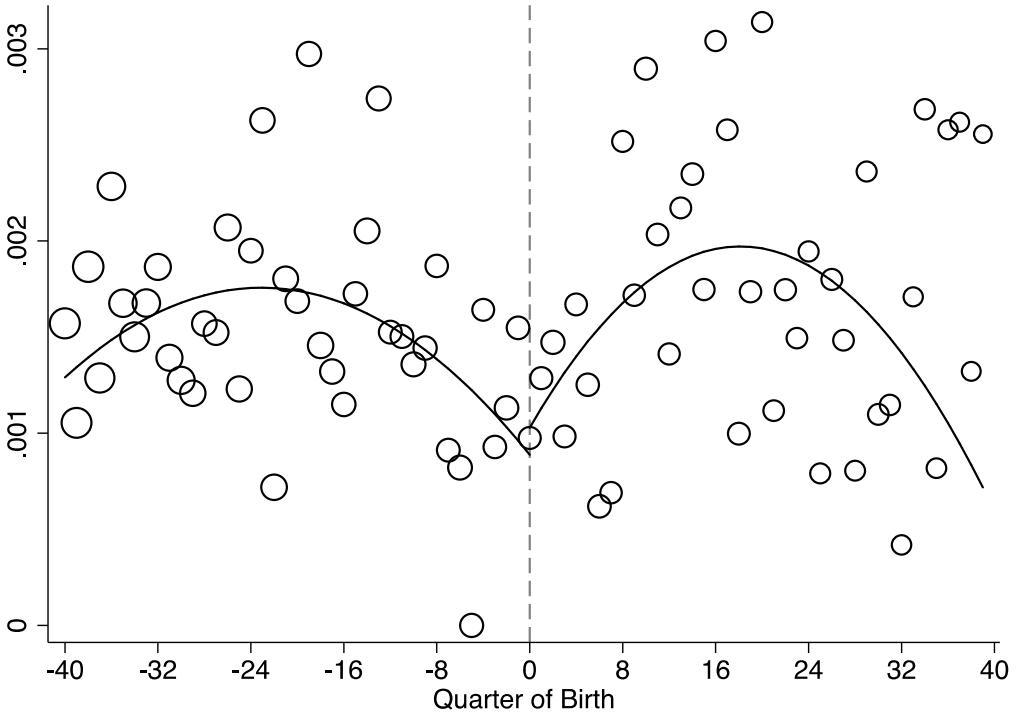
Notes: The figure shows the fraction of study participants for whom data on body size index was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while 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



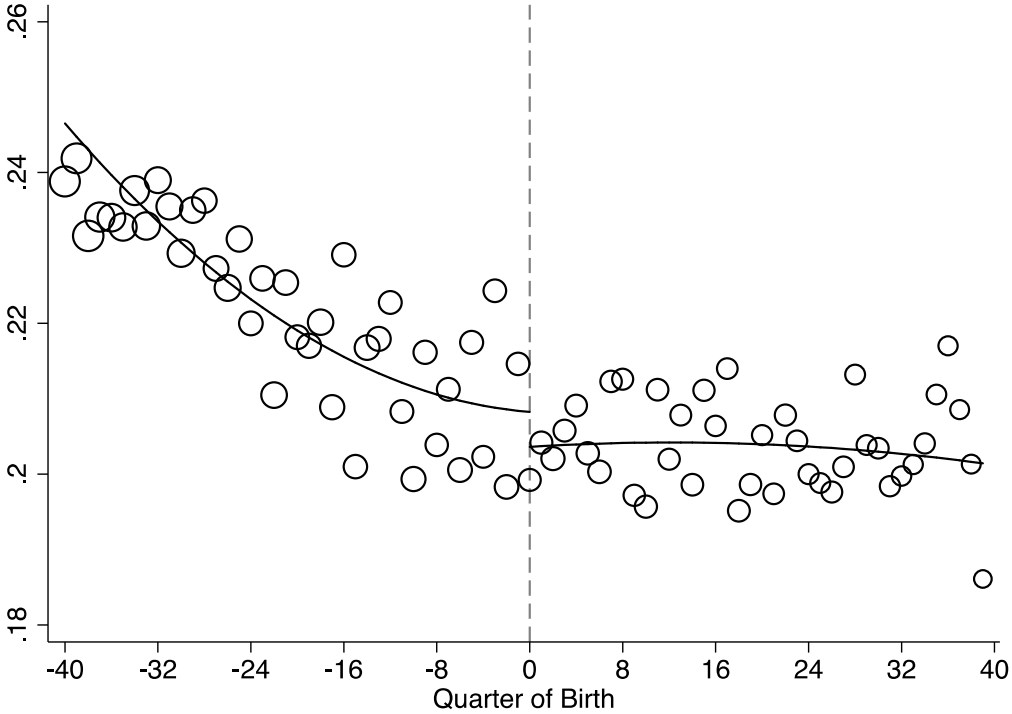
Notes: The figure shows the fraction of study participants for whom spirometry data was missing by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts 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

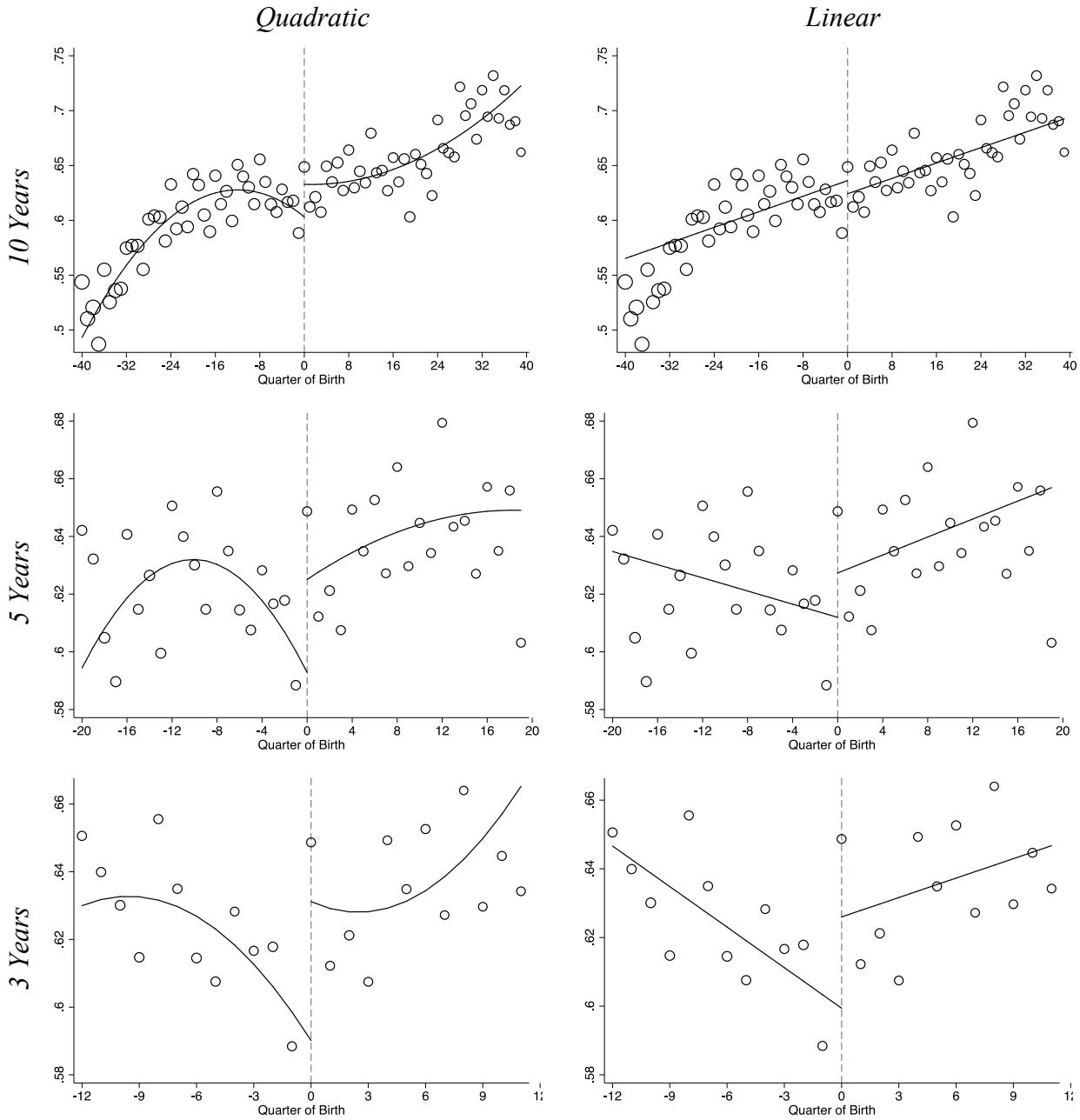
Notes: The table investigates whether there are discontinuities in missing outcomes at the September 1, 1957 cutoff. It reports the coefficient on an indicator for being born on or after September 1, 1957 (i.e., "Post") from regressions where the dependent variables is listed in the column. The regressions 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



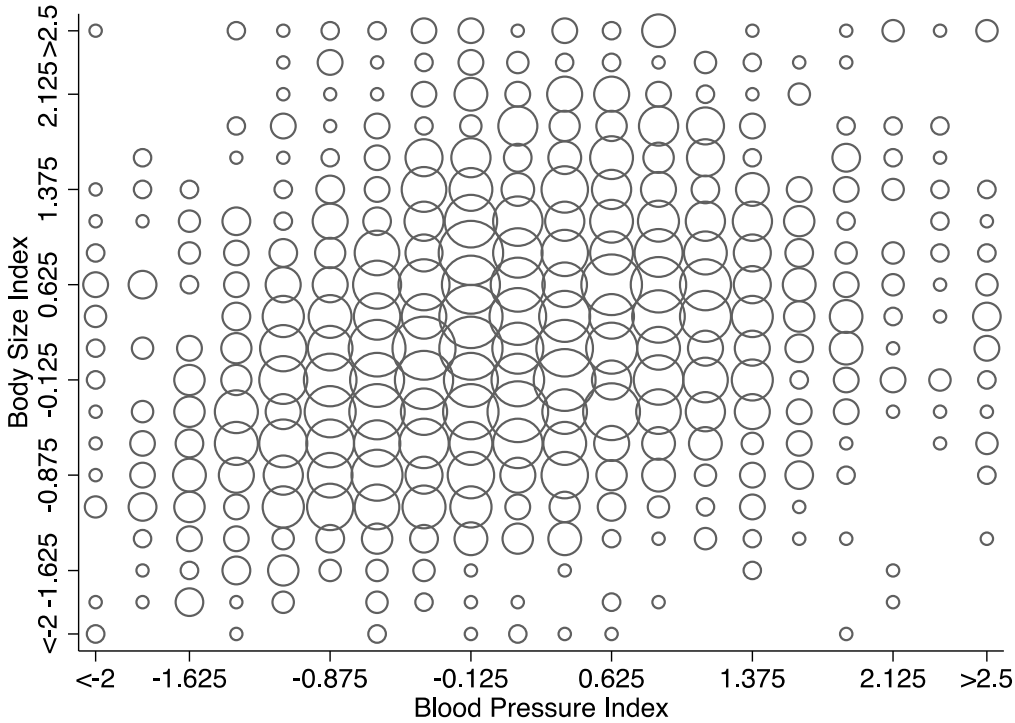
Notes: The figure shows the fraction of study participants who stayed in school until age 17 by quarter of birth for different specifications. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear trends in quarter of birth. The top row uses a 10-year 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



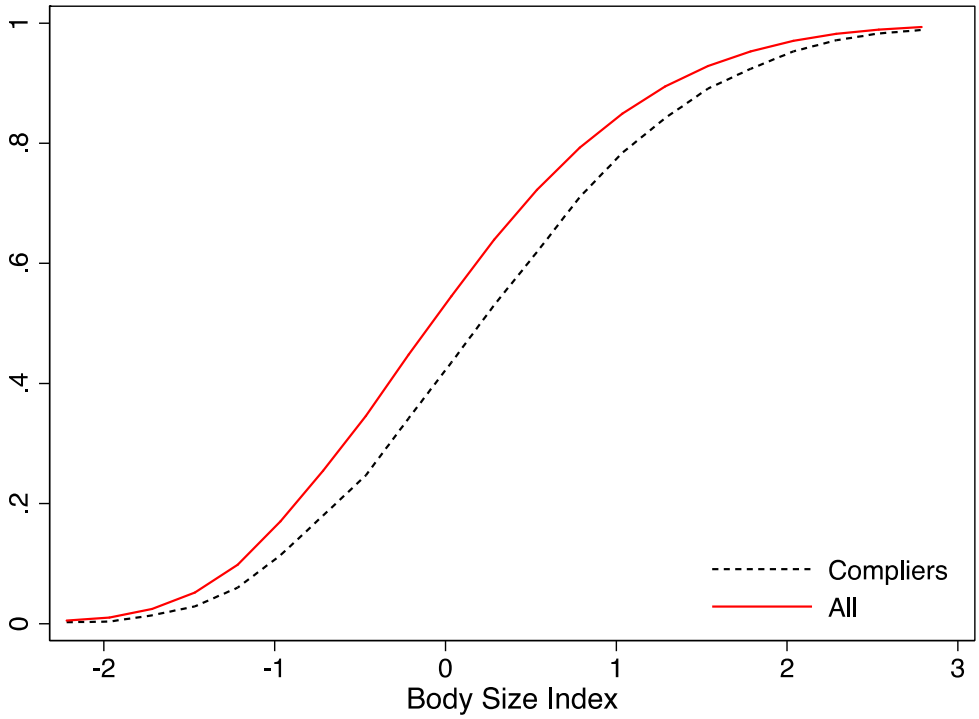
Notes: The figure shows the joint distribution of body size and blood pressure indices among compliers born in the 12 months before September 1, 1957. The circumference of each circle reflects the mass in that interval. $N = 2{,}210$.

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

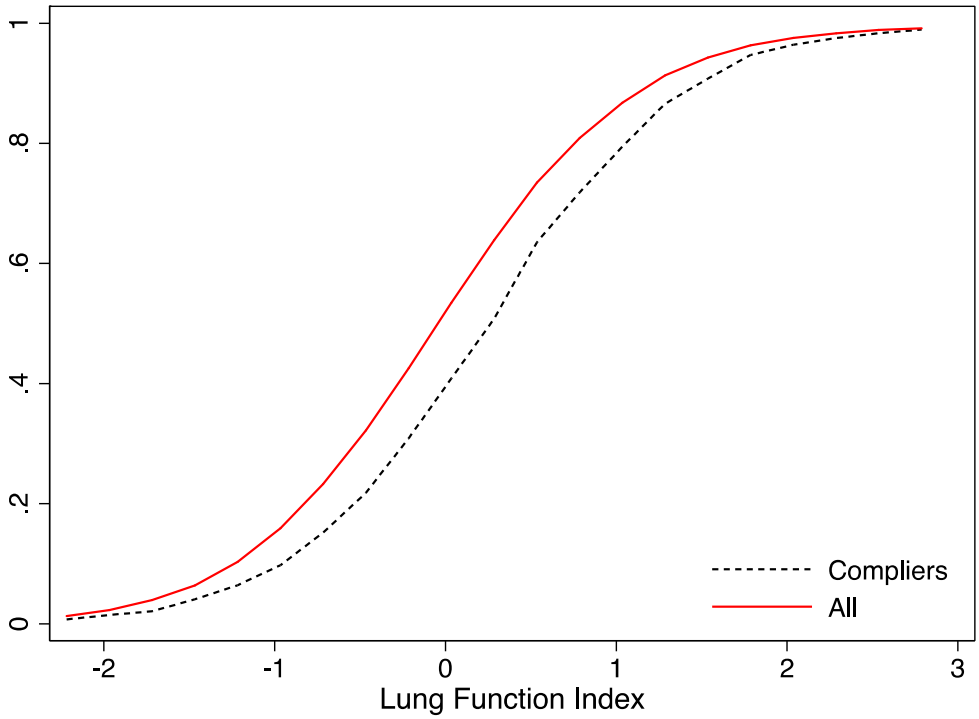
Notes: The table shows the cateogires of the 2000 National Statistics Socio-economic Classification (NS-SEC) of occupations.

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



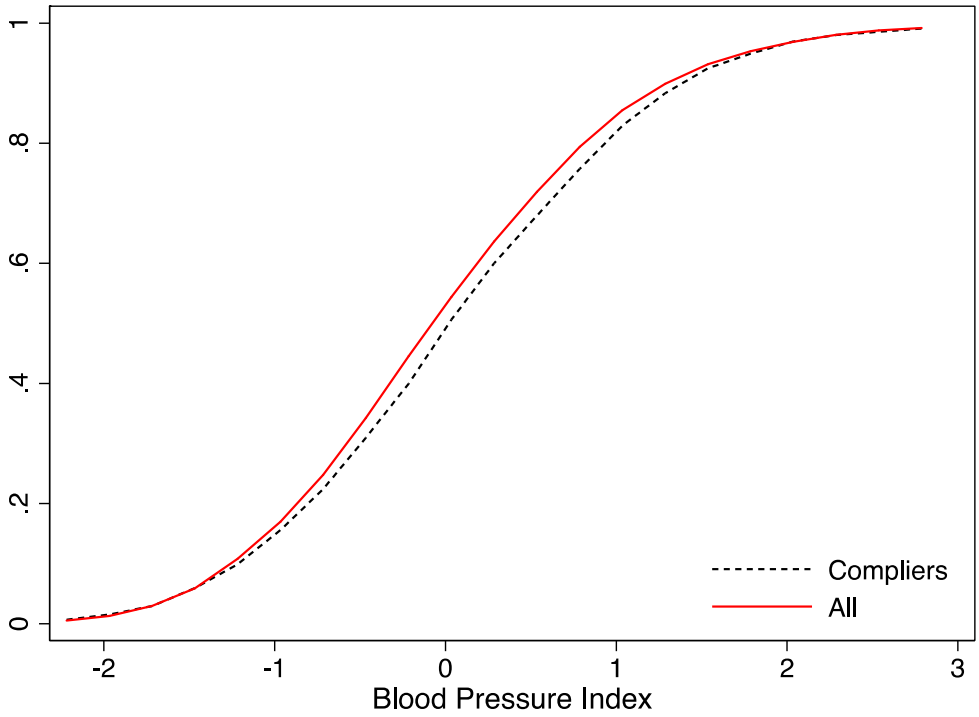
Notes: The figure shows the pre-reform CDFs of body size index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. N = 33,228 (compliers) and 158,707 (all).

Appendix Figure D5: Pre-Reform Cumulative Distribution of Lung Function Index for Compliers and for Entire Population



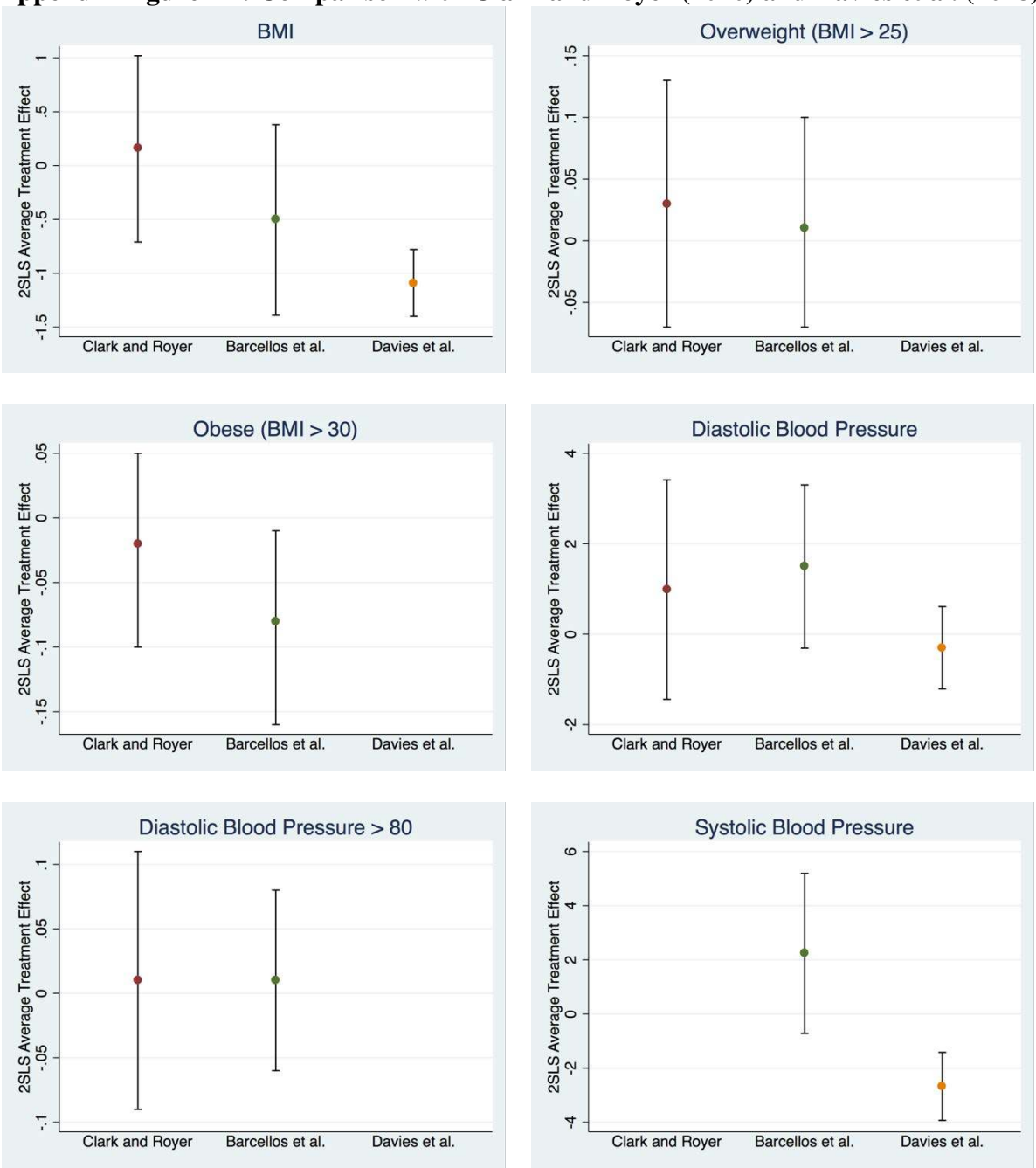
Notes: The figure shows the pre-reform CDFs of lung function index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. N = 25,021 (compliers) and 127,195 (all).

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).

Appendix Figure D7: Comparison with Clark and Royer (2010) and Davies et al. (2018)



Notes: The figures compare the point estimates and the 95% confidence intervals for 2SLS estimates of the effect of staying in school until age 16. See respective papers for details about bandwidth, controls, and trends.

Appendix Table D3: Comparison with Clark and Royer (2010) and Davies et al. (2018)

	Blood Pressure			*Body Mass Index*		
	Diastolic	*Diastolic > 90*	*Systolic*	*BMI*	*BMI > 25*	*BMI > 30*
Clark & Royer						
Point estimate	0.99	0.01		0.16	0.03	-0.02
Lower Bound 95% CI	-1.44	-0.09		-0.71	-0.07	-0.10
Upper Bound 95% CI	3.41	0.11		1.02	0.13	0.05
N	15,097	15,097		18,473	18,473	18,473
Barcellos et al.						
Point estimate	1.49	0.01	2.23	-0.50	0.01	-0.08
Lower Bound 95% CI	-0.31	-0.06	-0.72	-1.39	-0.07	-0.16
Upper Bound 95% CI	3.30	0.08	5.19	0.38	0.10	-0.01
N	270,647	270,647	270,647	269,970	269,970	269,970
Davies et al.						
Point estimate	-0.30		-2.67	-1.09		
Lower Bound 95% CI	-1.21		-3.93	-1.40		
Upper Bound 95% CI	0.61		-1.42	-0.78		
N	21,494		21,492	22,055		

Notes: The table compares the point estimates and the 95% confidence intervals for 2SLS estimates of the effect of staying in school until age 16 on the health outcomes shown in the columns. See respective papers for details about bandwidth, controls, and trends.

Appendix Table D4: Clustering Standard Errors by Day-Month-Year of Birth

	Left school at age ≥ 16				*CSE or -level*			
First stage								
Post	0.150	0.150	0.139	0.139	0.076	0.076	0.079	0.079
	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.006]***	[0.006]***	[0.006]***	[0.006]***
Clustered SEs?	No	Yes	No	Yes	No	Yes	No	Yes
Controls?	No	No	Yes	Yes	No	Yes	No	Yes
N Observations	271,082				268,551			
	Body Size				*Lung Function*			
Two stages least squares								
Stayed in school until 16	-0.154	-0.154	-0.163	-0.163	-0.175	-0.175	-0.174	-0.174
	[0.083]*	[0.087]*	[0.091]*	[0.094]*	[0.103]*	[0.105]*	[0.112]	[0.114]
Clustered SEs?	No	Yes	No	Yes	No	Yes	No	Yes
Controls?	No	No	Yes	Yes	No	Yes	No	Yes
N Observations	266,525				215,536			
	Blood Pressure				*Summary*			
Two stages least squares								
Stayed in school until 16	0.151	0.151	0.151	0.151	-0.120	-0.120	-0.125	-0.125
	[0.084]*	[0.086]*	[0.091]*	[0.093]	[0.103]	[0.105]	[0.112]	[0.115]
Clustered SEs?	No	Yes	No	Yes	No	Yes	No	Yes
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
N Observations	270,647				212,689			

Notes: The table shows how the standard error estimates change when we cluster the standard errors by day-month-year of birth.

Appendix E

Appendix Table E1: 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.

Appendix Table E2: 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. Lower values correspond to higher SES. *N* = 207,533.

Appendix Table E3: 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.

Appendix Table E4: Effect on Diet

	Calories	*% Sugars*	*% Fat*	*% Saturated Fat*	*% Carbo-hydrates*
Stayed in school until 16	-86.868	0.020	-0.030	-0.019	0.011
	[153.979]	[0.017]	[0.017]*	[0.008]**	[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.

Appendix Table E5:
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.

Appendix F

Power to Detect Mean Effects Versus Distributional Effects

Introduction

In this section, we illustrate a case where an estimate of average treatment effects will be lower powered than distributional treatment effects. More precisely, we examine a setting where only individuals in the upper portion of the outcome distribution are responsive to treatment. We calculate analytically the power to estimate the treatment effect on the average value of the outcome in the population.

Due to the analytic and computation complexity of the Anderson-Darling-based test used in this paper, we are unable to calculate the power of that test analytically nor by simulation. Instead, we evaluate the power to measure the treatment effect on an indicator of whether the outcome variable is above or below certain values at different parts of the outcome distribution. (See details below.) While this is not a perfect comparison, it is meant to give intuition for when distributional tests may perform better than tests of the average.

We find in this simplified setting that when most individuals are affected by some treatment, estimates of average effects are better powered than estimates of distributional effects. However, when fewer individuals are affected, distributional tests may be better powered.

Data Generating Process

Let Y_i denote some outcome of interest for individual i. We are interested in the effect of some treatment X_i on Y_i, where X_i is an indicator variable of whether individual i was treated. Let $Y_{0,i}$ denote the potential outcome of individual i in the case that they were not treated and $Y_{1,i}$ denote the potential outcome of individual i in the case that they were.

To simulate distributional effects, we assume that only the top p_τ fraction of the potential outcome distribution is affected by treatment. More precisely, we assume that

$$Y_{1,i} = \begin{cases} Y_{0,i} & \text{if } Y_{0,i} \leq \tau \\ Y_{0,i} + \delta & \text{otherwise,} \end{cases}$$

where τ is the (1- p_τ)-th percentile of the potential outcome distribution, $\tau \equiv \Phi^{-1}(1\text{-}\, p_\tau)$, and δ is the effect of treatment on those affected. To maintain monotonicity and simplify this derivation, we assume that $\delta > 0$.

In this exercise, we assume that we draw a sample of N individuals from the population, treat a fraction p_x of them, and measure their realized outcome Y_i. In the sections below, we calculate the power to find a statistically significant effect of treatment by looking at the effect of treatment on the average of the outcome and by looking at the effect of treatment on specific parts of the outcome distribution.

The Distributional Test

Because the distributional test used in this paper is very complicated, both analytically and computationally, it will not be possible to calculate the power of that test. We instead calculate the power of an alternative but related test that is meant to provide intuition for why and under which circumstances a distributional test may be better powered than a test of an average treatment effect.

Specifically, we will consider an indicator variable, T_i, for whether Y_i is greater than some value t, and calculate the power to detect an effect of the treatment on T_i. Recall that the distributional test in this paper is a weighted integral of these treatments effects across a range of values of t. Thus, if for the values of t in that range, the power of each corresponding test is greater than the power to detect an effect on the average value of Y_i, it is likely that the power of the distributional test will similarly be greater.

Some Intermediate Calculations

In order to perform the power calculations below, we will need to know the values of $\mathrm{E}(Y_i)$, $\mathrm{Var}(Y_i)$, $\mathrm{E}(T_i)$, and $\mathrm{Var}(T_i)$. We first calculate

$$\mathrm{E}(Y_i) = \mathrm{P}(X_i = 0)\mathrm{E}(Y_i|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(Y_i|X_i = 1)$$

$$= \mathrm{P}(X_i = 1)\mathrm{E}(Y_i|X_i = 1)$$

$$= p_x p_\tau \delta.$$

Next, we calculate

$$\mathrm{E}(Y_i^2) = \mathrm{P}(X_i = 0)\mathrm{E}(Y_i^2|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(Y_i^2|X_i = 1)$$

$$= (1 - p_x) + p_x\mathrm{E}(Y_i^2|X_i = 1).$$

Note that the variable $(Y_i \mid X_i = 1)$ is the same as $(Z + \delta I)$ where Z is a standard normal random variable and I is an indicator variable for whether $Z > \tau$. We therefore continue

$$\mathrm{E}(Y_i^2) = (1 - p_x) + p_x\mathrm{E}[(Z + \delta I)^2]$$

$$= (1 - p_x) + p_x[\mathrm{E}(Z^2) + 2\mathrm{E}(ZI)\delta + \mathrm{E}(I^2)\delta^2]$$

$$= (1 - p_x) + p_x[1 + 2\phi(\tau)\delta + p_\tau\delta^2]$$

$$= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau \delta^2,$$

where $\phi(.)$ is the standard normal pdf. Finally, this implies that

$$\mathrm{Var}(Y_i) = \mathrm{E}(Y_i^2) - \mathrm{E}(Y_i)^2$$

$$= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau \delta^2 - (p_x p_\tau \delta)^2$$

$$= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau(1 - p_x p_\tau)\delta^2.$$

For the binary variable, we first calculate

$$\mathrm{E}(T_i) = \mathrm{P}(X_i = 0)\mathrm{E}(T_i|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(T_i|X_i = 1)$$

$$= (1 - p_x)[1 - \Phi(t)] + p_x[1 - \Phi(t - \delta)]$$

$$= 1 - (1 - p_x)\Phi(t) - p_x\Phi(t - \delta).$$

This implies that

$$\mathrm{Var}(T_i) = [(1-p_x)\Phi(t) + p_x\Phi(t-\delta)][1-(1-p_x)\Phi(t) - p_x\Phi(t-\delta)].$$

Power to Detect Changes in the Average Outcome

To estimate the effect of treatment on average health, β_{avg}, we regress Y_i on X_i. In that case, we see that

$$\beta_{\text{avg}} = \frac{\mathrm{Cov}(X_i, Y_i)}{\mathrm{Var}(X_i)}$$

$$= \frac{\mathrm{E}(X_i Y_i) - \mathrm{E}(X_i)\mathrm{E}(Y_i)}{\mathrm{Var}(X_i)}$$

$$= \frac{p_\tau \delta p_x - p_\tau \delta p_x^2}{p_x(1-p_x)}$$

$$= p_\tau \delta.$$

Using the derivations from the previous section, the standard error an estimator of β_{avg} is

$$\mathrm{SE}(\beta_{\text{avg}}) = \sqrt{\frac{\mathrm{Var}(Y_i) - \mathrm{Var}(X_i\beta_{\text{avg}})}{N\,\mathrm{Var}(X_i)}}$$

$$= \sqrt{\frac{1 + 2p_x\phi(\tau)\delta + p_x p_\tau(1-p_\tau)\delta^2}{N\,p_x(1-p_x)}}.$$

From these expressions, we see that the z-statistic for the average effect will be distributed as

$$N\left(p_\tau\delta\sqrt{\frac{N\,p_x(1-p_x)}{1 + 2p_x\phi(\tau)\delta + p_x p_\tau(1-p_\tau)\delta^2}}, 1\right).$$

So the power of a test of whether there is a non-zero mean effect of the policy will be equal to the fraction of the time that this normally distributed random variable achieves a value greater than 1.96 in magnitude.

Power to Detect Changes in the Distribution of the Outcome

Let T_i be defined for some threshold t, as described above. We first note that if $t < \tau$, then the treatment effect, β_t, will be equal to zero since the treatment will not induce any individuals below the threshold to cross the threshold. This means that power to estimate an effect of treatment with a p-value of less than 0.05 is 5%.

We next consider that case that $t > \tau + \delta$. In this setting, every individual less than δ units below the threshold t will be above the threshold after treatment. Therefore

$$\beta_t = E(T_i|X_i = 1) - E(T_i|X_i = 0)$$

$$= E(Y_i > t|X_i = 1) - E(Y_i > t|X_i = 0)$$

$$= 1 - \Phi(t - \delta) - 1 + \Phi(t)$$

$$= \Phi(t) - \Phi(t - \delta)$$

$$= p_t - p_{t-\delta},$$

where $p_t \equiv \Phi(t)$ is the fraction of individuals where $Y_{0,i} \leq t$.

The standard error is therefore

$$\mathrm{SE}(\beta_t) = \sqrt{\frac{\mathrm{Var}(T_i) - \mathrm{Var}(X_i\beta_t)}{N\,\mathrm{Var}(X_i)}}$$

$$= \sqrt{\frac{(1 - p_x)p_t(1 - p_t) + p_x p_{t-\delta}(1 - p_{t-\delta})}{N\,p_x(1 - p_x)}}.$$

This means the z-statistics for a test of a treatment effect on T_i is distributed

$$N\left((p_t - p_{t-\delta})\sqrt{\frac{N\,p_x(1 - p_x)}{(1 - p_x)p_t(1 - p_t) + p_x p_{t-\delta}(1 - p_{t-\delta})}}, 1\right).$$

As with the calculation of power in the average treatment effect-case, the power in this setting is the fraction of time that this normally distributed random variable exceeds one in absolute value.

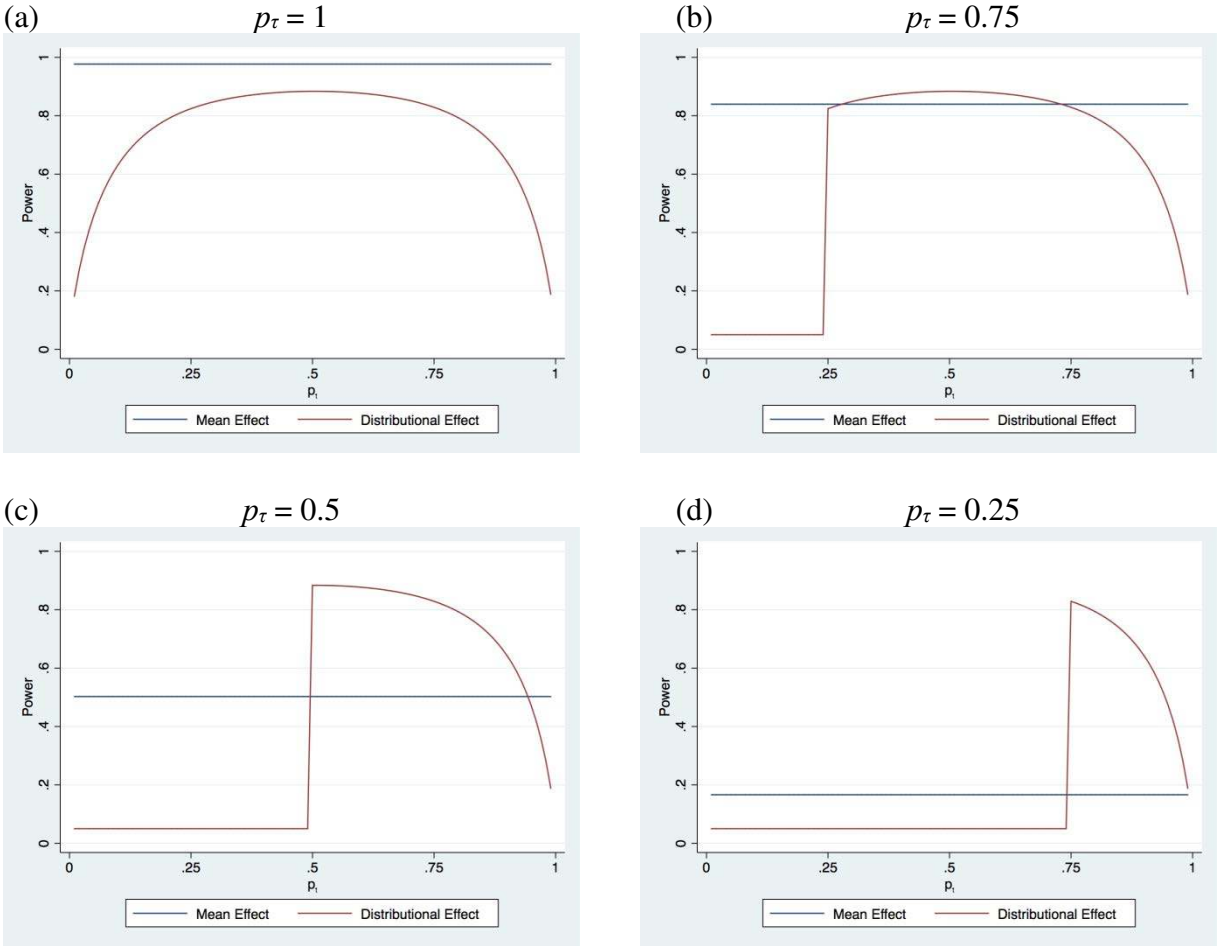
We do not calculate the power of a test for values of t in $[\tau, \tau + \delta]$ but rather note that the power will be somewhere between the cases when t is above or below this interval.

Illustration of Power Calculations in Various Settings

In this illustration, we compare the power of the estimates of the treatment on the average outcome and the effect of the treatment on the binary outcomes. In these calculations, we set $\delta = 0.025$, which is approximately the same magnitude as the estimated average effect for all three health indices in this paper. Qualitatively, the results of this illustration are the same at any sample size, but we set $N = 100{,}000$ here because it makes the results easier to display. We set $p_x = 0.5$.

In order to investigate our claim that distributional effects may be better powered when not everyone is affected homogeneously by the treatment, we consider the cases $p_\tau \in \{0.25, 0.5, 0.75, 1\}$. We evaluate the power for the continuous and binary case for a dense range of values of $p_t \in (0, 1)$. The results of these calculations are found in Appendix Figure 1 below.

Appendix Figure F1: Power of the Mean vs Distributional Effect Estimates



Discussion

Panel (a) corresponds to a setting when every individual is affected by the treatment homogeneously. Unsurprisingly, the power of the average effect estimate is greater than the power of the distributional effect estimate for all values of p_t.

Panel (b) corresponds to a setting where only 25% of individuals are unaffected by the policy. Note that for values of p_t less than 0.25, the distributional effect has very low power, a result of individuals in that part of the distribution being unaffected by the treatment. Nevertheless, over a large interval of values for p_t, the distributional test is slightly better powered than the test on the average. The power of the distributional test quickly decays outside of this range however, suggesting that a test that considers all values of p_t in the range 0.5 to 1 (e.g., the Anderson-Darling test used in this paper) may not be better powered than a test on the average.

In panel (c), however, where 50% of individuals are unaffected, the difference in power becomes substantial over nearly the whole range of values in the upper half of the distribution. This is particularly relevant to our setting because, observing the pre- and post-reform CDF of body size index, it appears that the ROSLA affected only for those in the upper half of the distribution.

The patterns in paned (d), with 75% of the population unaffected, are similar to those of the first three panels. For the values of p_t corresponding to affected individuals, the difference in power

between the distributional and average effect is even larger, though the range of individuals who are affected is much narrower than in the other panels.

This discussion shows how tests of distributional effects may be better powered than tests for average effects in certain cases. Obviously, this framework is simplified in order to make the math tractable, but the general principle will hold that when only a portion of the distribution is affected by some treatment, methods that focus on those segments of the distribution may be better powered than those that consider the whole distribution. This appears to be increasingly true as the fraction affected becomes smaller.

Appendix G

We conducted a back-of-envelope calculation to estimate the mortality consequences of the estimated reduction in BMI caused by staying in school until age 16. To map BMI into mortality, we used Aune et al. (2016)'s estimates of the association between BMI and all cause mortality. In particular, we used the estimates from the column "All participants" of Table 2.

In the first step, we fitted a fractional polynomial of second order through the points in Table 2 in order to obtain a continuous mapping of relative mortality as a function of BMI. A fractional polynomial is a polynomial that may include logarithms, noninteger powers, and repeated powers. Here is the Stata output from this estimation:

```
.         fp <BMI>: regress RR <BMI>
(fitting 44 models)
(....10%....20%....30%....40%....50%....60%....70%....80%....90%....100%)

Fractional polynomial comparisons:
```

BMI	df	Deviance	Res. s.d.	Dev. dif.	P(*)	Powers
omitted	0	22.177	0.500	76.310	0.000	
linear	1	20.454	0.490	74.588	0.000	1
m = 1	2	16.741	0.436	70.875	0.000	3
m = 2	4	-54.133	0.049	0.000	--	-2 3

(*) P = sig. level of model with m = 2 based on F with 11 denominator dof.

Source	SS	df	MS
Model	3.71445463	2	1.85722731
Residual	.031789143	13	.002445319
Total	3.74624377	15	.249749585

Number of obs	=	16
F(2, 13)	=	759.50
Prob > F	=	0.0000
R-squared	=	0.9915
Adj R-squared	=	0.9902
Root MSE	=	.04945

RR	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
BMI_1	540.4458	16.46243	32.83	0.000	504.8809	576.0107
BMI_2	.0000281	7.26e-07	38.70	0.000	.0000265	.0000297
_cons	-.3610887	.0493578	-7.32	0.000	-.4677197	-.2544577

where RR is the relative mortality and $BMI_1 = BMI^{-2}$ and $BMI_2 = BMI^{3}$.

Let $\widehat{RR}(x)$ be the predicted relative risk of death for an individual with a BMI of x.

In the second step, we estimated the vintiles of the distribution of BMI among the compliers born between September 1, 1956 and August 31, 1957, where compliers are defined as those born before the reform who dropped out of school at age 15 and younger. These vintiles define the grid used to estimate the pre- and the post-reform cumulative distribution functions of BMI.

In the third step, we estimated $F_{pre}(\tau)$ and $F_{post}(\tau)$ at each vintile τ (following the procedure explained in Section 3 of the paper).

In the fourth step estimated, we estimated the pre- and post-reform relative risk of death:

$$\widehat{Risk}_{pre} = \widehat{RR}(\tau_5) * \hat{F}_{pre}(\tau_5) + \sum_{k=10,15,\ldots,90} \widehat{RR}(\tau_k) * \left[\hat{F}_{pre}(\tau_k) - \hat{F}_{pre}(\tau_{k-5})\right] + \\ + \widehat{RR}(\tau_{95}) * (1 - \hat{F}_{pre}(\tau_{90}))$$

$$\widehat{Risk}_{post} = \widehat{RR}(\tau_5) * \hat{F}_{post}(\tau_5) + \sum_{k=10,15,\ldots,90} \widehat{RR}(\tau_k) * \left[\hat{F}_{post}(\tau_k) - \hat{F}_{post}(\tau_{k-5})\right] + \\ + \widehat{RR}(\tau_{95}) * (1 - \hat{F}_{post}(\tau_{90}))$$

$\left[\left(\frac{\widehat{Risk}_{post}}{\widehat{Risk}_{pre}}\right) - 1\right] * 100$ is the estimated reduction in mortality associated with our distributional treatment effects.

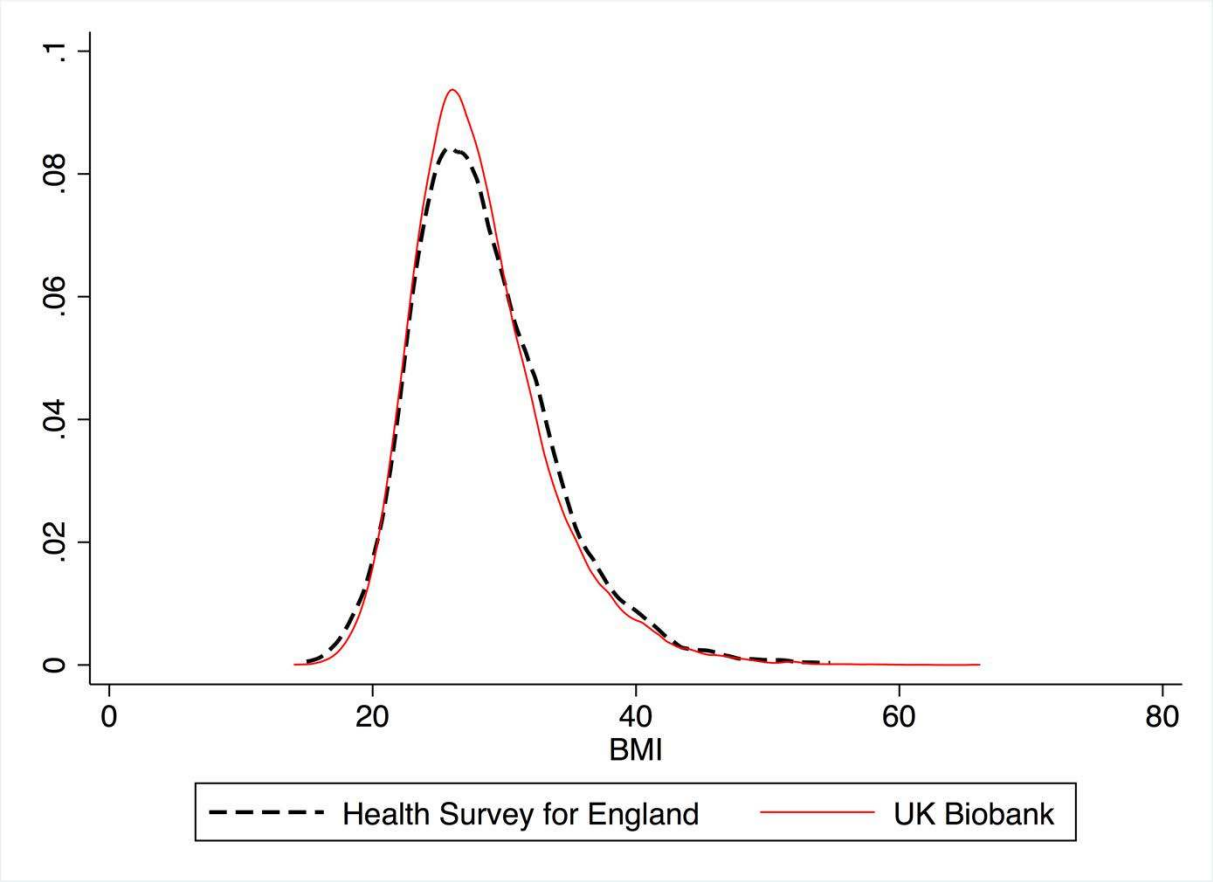
To calculate the reduction in mortality implied by the average treatment effect, we first estimated the average BMI for compliers born at September 1, 1957. To do that, we restricted the sample to participants born before September 1, 1957 who dropped out of school at age 15 or younger and run a regression of BMI on quadratic trends for date of birth. The coefficient on the constant, 28.51707, is our estimate of the pre-reform average BMI among compliers. Next, we estimated the 2SLS effect of staying in school until age on BMI, -.4066152. The post-reform average BMI among compliers is equal to 28.1104548 = 28.51707-.4066152. Finally, the estimated reduction in mortality is $\left[\left(\frac{\widehat{RR}(28.1104548)}{\widehat{RR}(28.51707)}\right) - 1\right] * 100$

REFERENCE

Aune, D., Sen, A., Prasad, M., Norat, T., Janszky, I., Tonstad, S., Romundstad, P. and Vatten, L.J., 2016. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *Bmj*, 353, p.i2156.

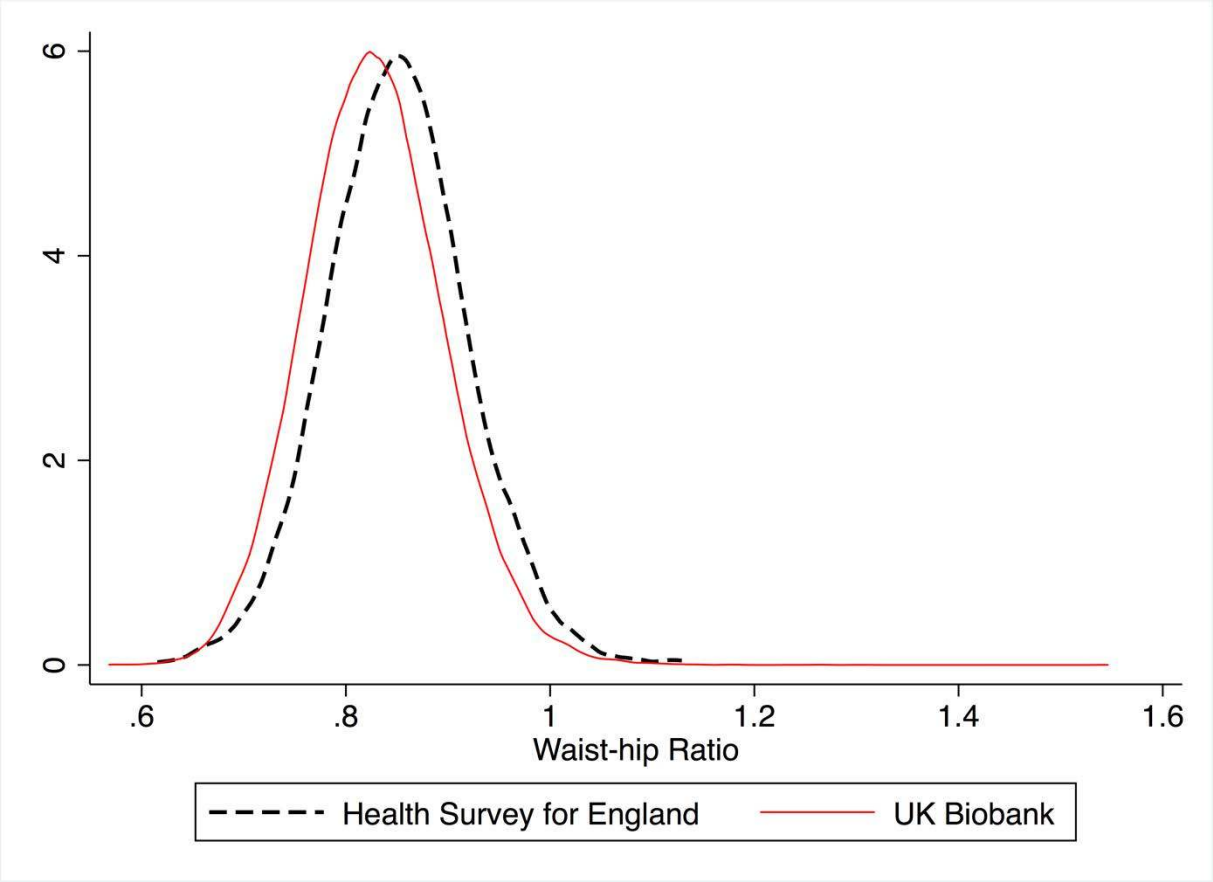
Appendix H

Appendix Figure H1: Comparison of Distribution of BMI of Compliers in the UK Biobank and in the Health Survey for England



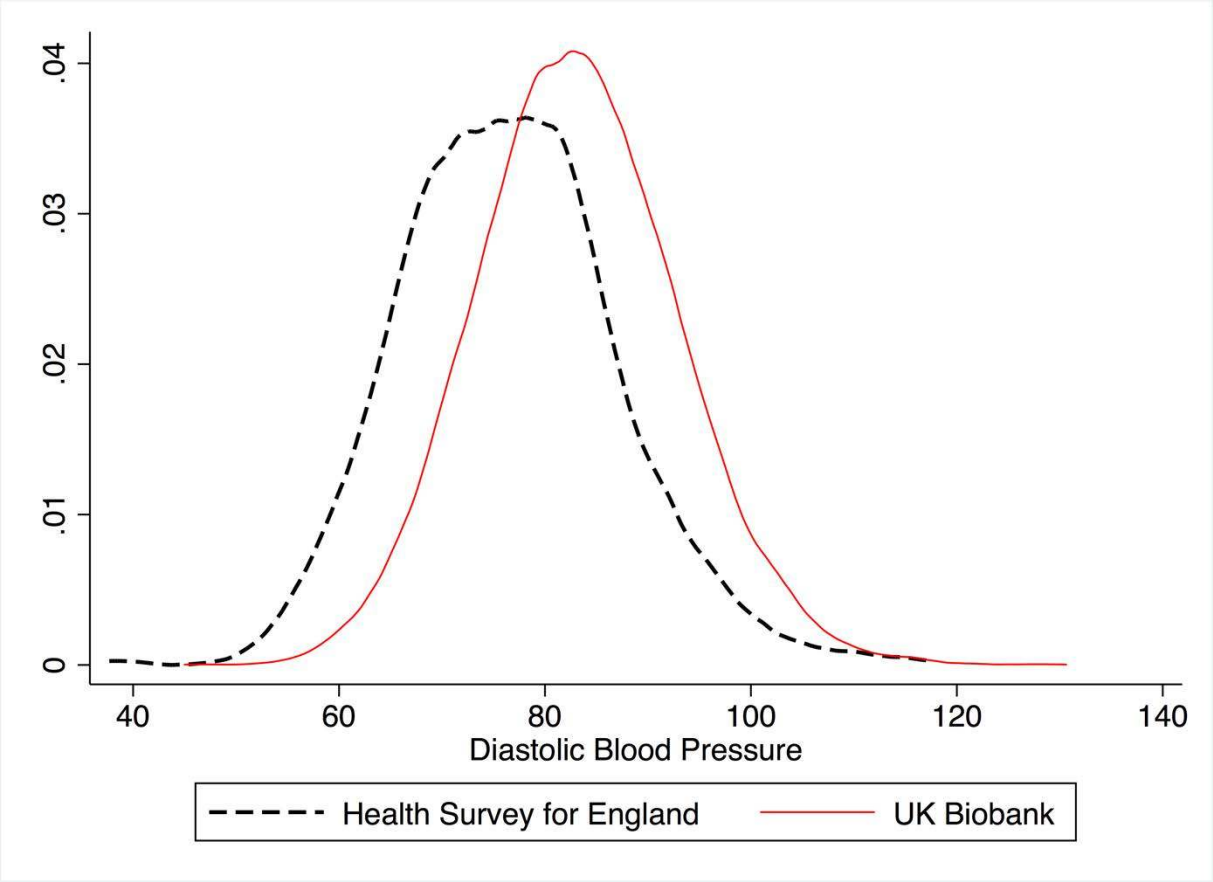
Notes: The figure compares the distribution of body mass index (BMI) of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents so we had to rely on age and month of interview to identify respondents who were born unambiguously before September 1957. We applied the same sample restrictions in terms of age (and month of interview) to both samples. We approximate the population of compliers as those born before September 1957 who dropped out of school at age 15 or younger. The HSE estimates include sample weights. The distributions are adjusted for differences in gender and age (using the HSE as reference). N = 39,186 (UKB), 2,700 (HSE).

Appendix Figure H2: Comparison of Distribution of Waist-Hip Ratio of Compliers in the UK Biobank and in the Health Survey for England



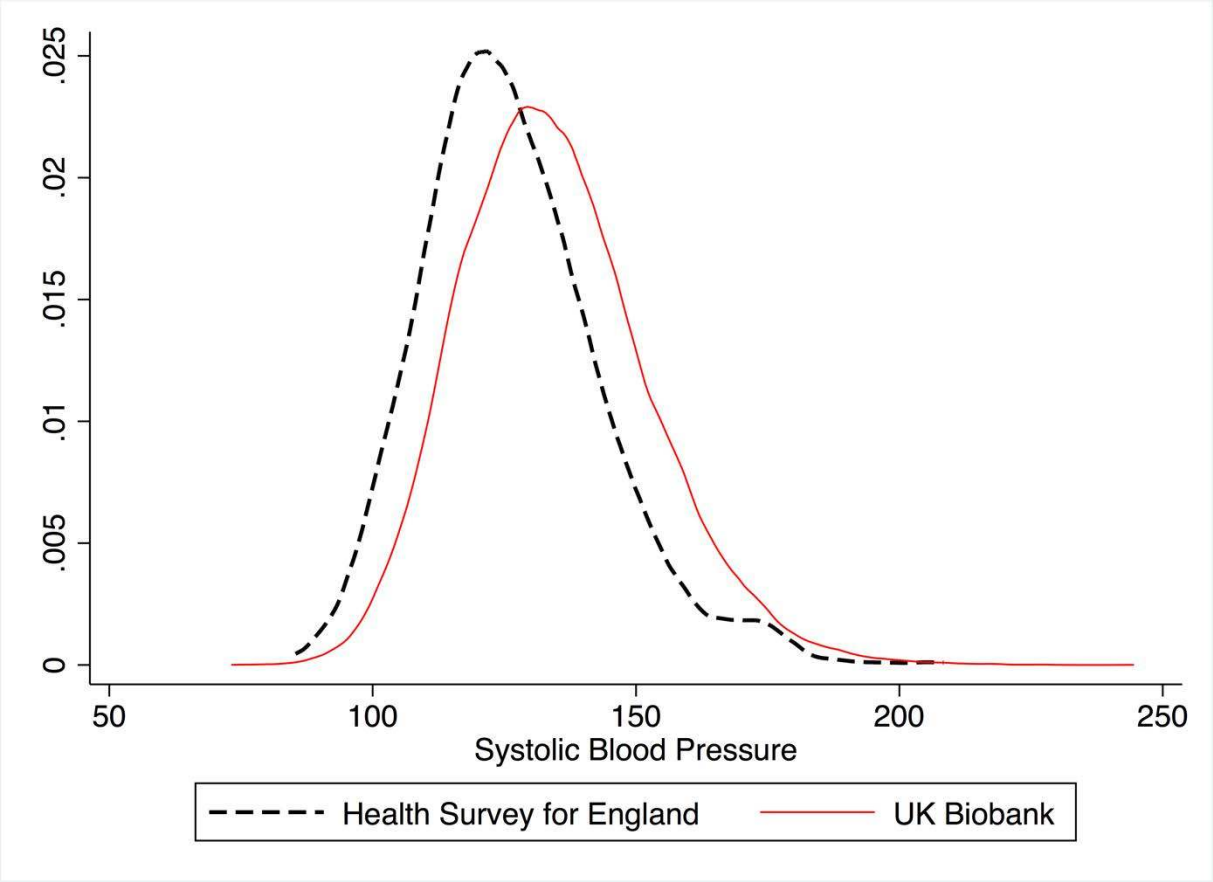
Notes: The figure compares the distribution of waist-hip ratio of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents so we had to rely on age and month of interview to identify respondents who were born unambiguously before September 1957. We applied the same sample restrictions in terms of age (and month of interview) to both samples. We approximate the population of compliers as those born before September 1957 who dropped out of school at age 15 or younger. The HSE estimates include sample weights. The distributions are adjusted for differences in gender and age (using the HSE as reference). N = 39,249 (UKB), 2,259 (HSE).

Appendix Figure H3: Comparison of Distribution of Diastolic Blood Pressure of Compliers in the UK Biobank and in the Health Survey for England



Notes: The figure compares the distribution of diastolic blood pressure of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents so we had to rely on age and month of interview to identify respondents who were born unambiguously before September 1957. We applied the same sample restrictions in terms of age (and month of interview) to both samples. We approximate the population of compliers as those born before September 1957 who dropped out of school at age 15 or younger. The HSE estimates include sample weights. The distributions are adjusted for differences in gender and age (using the HSE as reference). N = 39,316 (UKB), 1,899 (HSE).

Appendix Figure H4: Comparison of Distribution of Systolic Blood Pressure of Compliers in the UK Biobank and in the Health Survey for England



Notes: The figure compares the distribution of systolic blood pressure of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents so we had to rely on age and month of interview to identify respondents who were born unambiguously before September 1957. We applied the same sample restrictions in terms of age (and month of interview) to both samples. We approximate the population of compliers as those born before September 1957 who dropped out of school at age 15 or younger. The HSE estimates include sample weights. The distributions are adjusted for differences in gender and age (using the HSE as reference). N = 39,316 (UKB), 1,899 (HSE).

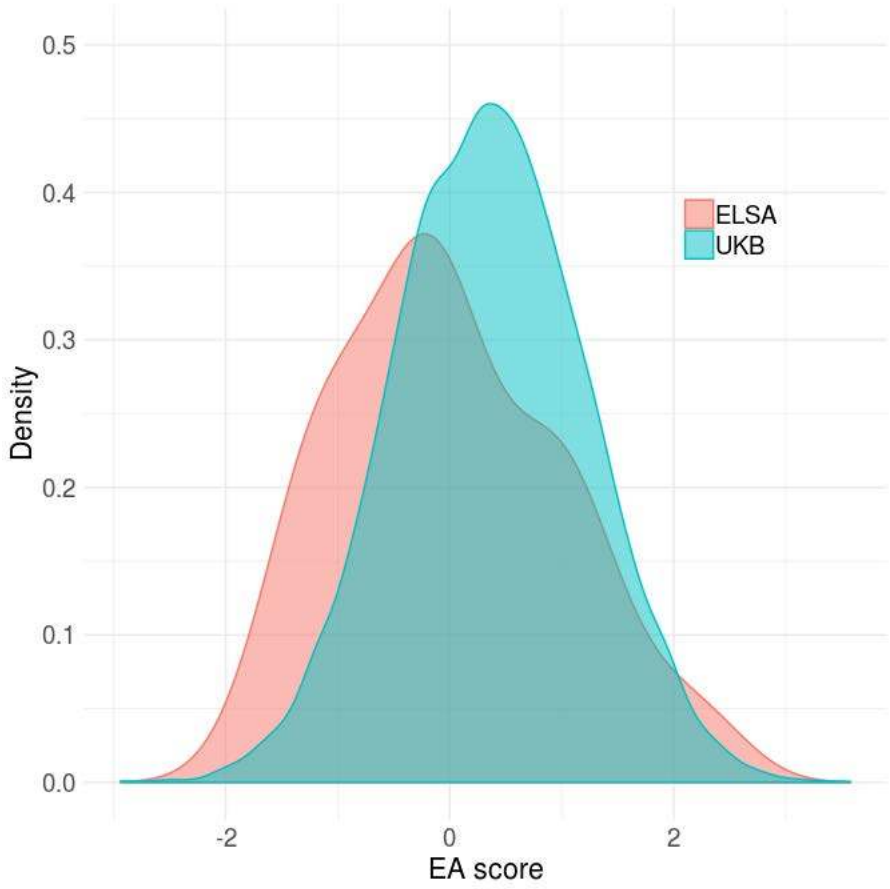
We used genetic data to re-weight the UK Biobank (UKB) sample in an attempt to make it nationally representative. Genetic data may be useful in this regard because it is fixed at conception. We used the English Longitudinal Survey of Ageing (ELSA) as a nationally representative benchmark. We restricted both samples to individuals of European ancestry in each data set who were born in the years 1954-1959.

We use a polygenic score (PGS) for educational attainment (EA) to compare the two samples. A PGS is a weighted sum of molecular genetic markers called single-nucleotide polymorphisms (SNPs). The weights for this PGS are based on a genome-wide association study (GWAS) of EA (Lee et al. 2018) and are derived using a standard procedure, LDpred (Vilhjálmsson et al. 2015). A PGS for EA constructed in this way reduces dimensionality of the genetic data while still maximizing its predictive power. This PGS has been shown to explain 11-13% of the variation in EA, slightly less than the variation explained by a parent's EA (Lee et al. 2018).

To compare the distribution of the EA PGS between the ELSA and UKB samples, it is important to construct the polygenic score from the same set of SNPs. The ELSA sample contains imputed genotypes for 39,131,557 SNPs. The UKB sample contains 16,642,384 imputed SNPs. In order to include only those SNPs that are well-imputed, we restrict SNPs in the ELSA sample with Rsq values of exactly 1, where Rsp is a measure of imputation quality between 0 and 1. Due to computational constraints, we were unable to include an imputation quality criterion on the UKB sample. Keeping only SNPs that are in the intersection of the UKB sample, ELSA sample, and the GWAS summary statistics, we are left with 281,700 SNPs. The PGS is constructed based on this set of SNPs.

Appendix Figure H5 shows the distribution of the polygenic score in the two samples.

Appendix Figure H5: Distribution of Polygenic Scores for Educational Achievement in the UK Biobank (UKB) and in the English Longitudinal Survey of Ageing (ELSA)



Notes: PGSs are standardized using the mean and standard deviation of PGSs in the ELSA sample.

To create weights, the distribution of scores is divided into 512 evenly-spaced points. At each point, the weight equals the ratio of the density of UKB observations to the density of ELSA observations. Then, for each individual in the UKB, we find the score distribution point nearest to the individual's score, and assign the weight associated with this point as the individual's weight.

Appendix Table H1 shows the results with and without weighting.

Appendix Table H1: Sensitivity of the Estimates of the Effects on Education to Re-weighting Based on Genetic Data

	Left school at age ≥ 16			
Post	0.15	0.17	0.14	0.16
	[0.004]***	[0.006]***	[0.004]***	[0.006]***
Weighted?	No	Yes	No	Yes
Controls?	No	No	Yes	Yes
Mean of Y	0.827			

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 = 264,066. The estimates in the first and third columns are slightly different from the estimates in Table 1 in the paper because the sample here is restricted to those of European ancestry in the UK Biobank who were genotyped.

Vilhjálmsson, B.J., J. Yang, H.K. Finucane, A. Gusev, S. Lindström, S. Ripke, ... & T. Hayeck. (2015). "Modeling linkage disequilibrium increases accuracy of polygenic risk scores." *The American Journal of Human Genetics*, **97**(4), 576-592.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, ..., David I. Laibson, Jian Yang, Magnus Johannesson, Philipp D. Koellinger, Patrick Turley, Peter M. Visscher, Daniel J. Benjamin, and David Cesarini (2018). "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics*, **50**, 1112-1121.

Appendix I

Implications of Homogeneous Effects on Distribution Effect Estimates

In this section, we will show that distributional effects behave predictably if there are homogeneous treatment effects. We begin by defining "distributional effects" with respect to the potential outcome distributions of compliers. For some outcome, y, let $F_{c,0}(y)$ denote the potential-outcome cumulative distribution function (CDF) of y for the set of compliers absent the policy, and let $F_{c,1}(y)$ denote the potential-outcome CDF of y for the set of compliers with the policy. If compliers all get exactly one additional year of school as a result of the policy, the distributional effect of an additional year of school in compliers is

$$\beta(y) = F_{c,1}(y) - F_{c,0}(y).$$

Notably, homogeneous treatment effects of a policy does not imply that for all y, $\beta(y)$ is constant. Rather, they imply that $\beta(y)$ is largest in parts of the distribution where y is most concentrated. This is because if every person is affected equally by some policy, more people will cross some threshold y_0 if there are more people near the y_0 threshold.

More formally, if a policy affects every individual by some constant, μ. Then $F_{c,1}(y) = F_{c,0}(y - \mu)$. This means that

$$\begin{aligned}\beta(y) &= F_{c,1}(y) - F_{c,0}(y)\\ &= F_{c,0}(y - \mu) - F_{c,0}(y).\end{aligned}$$

Recall that the definition of a derivative is

$$f'(x) = \lim_{\varepsilon \to 0} \frac{f(x) - f(x + \mu)}{\varepsilon}.$$

So for small μ, we have

$$\begin{aligned}\beta(y) &= F_{c,0}(y - \mu) - F_{c,0}(y)\\ &= \mu \frac{F_{c,0}(y - \mu) - F_{c,0}(y)}{\mu}\\ &= (-\mu) \frac{F_{c,0}(y) - F_{c,0}[y + (-\mu)]}{\mu}\\ &\approx -\mu\, F'_{c,0}(y)\\ &= -\mu\, f_{c,0}(y),\end{aligned}$$

where $f_{c,0}(y) \equiv F'_{c,0}(y)$ is the probability density function of y in compliers absent the policy. So we see that, if there are homogeneous treatment effects, the distributional effects would be approximately proportionate to the distribution of compliers.

Appendix J

Robustness to Extensive-Margin Heterogeneity in the First Stage

In this section we will show that our two-stage least squares estimates of distributional effects are not biased by extensive-margin heterogeneity in the first stage (see section 3 in the main paper for a definition of extensive- and intensive- margin heterogeneity in our context).

Imagine we are studying some policy that raises the schooling of a fraction of the population by exactly one year. Use π to denote the fraction of the population of "compliers." (i.e., those whose education is increased by a year when this policy is implemented.) We are interested in the effect of an additional year of schooling on some outcome, y (e.g., BMI), *in compliers*.

We begin by defining "distributional effects" with respect to the potential outcome distributions of compliers. Let $F_{c,0}(y)$ denote the potential-outcome cumulative distribution function (CDF) of y for the set of compliers absent the policy, and let $F_{c,1}(y)$ denote the potential-outcome CDF of y for the set of compliers with the policy. Since compliers all get exactly one additional year of school as a result of the policy, the distributional effect of an additional year of school *in compliers* is

$$\beta(y) = F_{c,1}(y) - F_{c,0}(y) .$$

The function $\beta(y)$ is the estimand of our two-stage least squares problem.

Next, we will show that under general conditions, two stage least squares obtains an estimate of $\beta(y)$, even if the first stage is heterogeneous in y. Let $F_{a,0}(y)$ denote the potential-outcome CDF of y for the set of always- and never-takers absent the policy, and let $F_{a,1}(y)$ denote the potential-outcome CDF of y for the set of always- and never-takers with the policy. By construction, always- and never-takers are not affected by the policy, and therefore $F_{a,0}(y) = F_{a,1}(y)$. Notably, we make no assumptions about the size of the first stage in different parts of the outcome distribution. It could be that there is a very large proportion of always-takers in the lower tail of the distribution of y and small proportion of always-takers in the upper tail, which would imply that the first stage is much bigger in the upper tail that the lower tail.

In practice, we cannot identify compliers in our data. We only can see the distribution of outcomes in the whole population. That is, without the policy, we observe the population outcome distribution

$$F_0(y) = \pi F_{c,0}(y) + (1 - \pi) F_{a,0}(y) .$$

With the policy, we observe the population outcome distribution

$$F_1(y) = \pi F_{c,1}(y) + (1 - \pi) F_{a,1}(y) .$$

In this framework, the First Stage is equal to the fraction of compliers in the entire population, $\beta_{FS} = \pi$. The Reduced Form is the difference in the outcome distribution with and without the policy in the whole population

$$\begin{aligned}
\beta_{RF}(y) &\equiv F_1(y) - F_0(y) \\
&= [\pi F_{c,1}(y) + (1 - \pi) F_{a,1}(y)] - [\pi F_{c,0}(y) + (1 - \pi) F_{a,0}(y)] \\
&= [\pi F_{c,1}(y) + (1 - \pi) F_{a,1}(y)] - [\pi F_{c,0}(y) + (1 - \pi) F_{a,1}(y)] \\
&= \pi [F_{c,1}(y) - F_{c,0}(y)] .
\end{aligned}$$

Our Two-Stage Least Squares estimate is then the ratio of the Reduced Form and First Stage estimates

$$\beta_{TSLS}(y) = \beta_{RF}(y) / \beta_{FS}$$
$$= \pi\,[F_{c,1}(y) - F_{c,0}(y)] / \pi$$
$$= F_{c,1}(y) - F_{c,0}(y)$$
$$= \beta(y),$$

which the distributional effect of an additional year of schooling *on compliers*. Therefore, two-stage least squares obtains true distributional-effect estimates even when there is a heterogeneous first stage.

Appendix K

Appendix Table K1: Effects of 1972 ROSLA on Education

	Left School at age ≥ 16		Left School at age ≥ 17		Left School at age ≥ 18		Left School at age ≥ 19	
Post	0.150	0.139	0.028	0.018	0.023	0.019	0.006	0.002
	[0.004]***	[0.004]***	[0.006]***	[0.006]***	[0.006]***	[0.006]***	[0.006]	[0.006]
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Observations	271,082	271,082	271,082	271,082	271,082	271,082	271,082	271,082
Mean of Y	0.827		0.612		0.532		0.442	

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.