Distributional Effects of Education on Health

Silvia H. Barcellos
Leandro S. Carvalho
Patrick Turley

Online Appendix

Appendix A

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, $\tau \equiv \Phi^{-1}(1- p_\tau)$, 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

$$\begin{aligned}\mathrm{E}(Y_i) &= \mathrm{P}(X_i = 0)\mathrm{E}(Y_i|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(Y_i|X_i = 1)\\ &= \mathrm{P}(X_i = 1)\mathrm{E}(Y_i|X_i = 1)\\ &= p_x p_\tau \delta.\end{aligned}$$

Next, we calculate

$$\begin{aligned}\mathrm{E}(Y_i^2) &= \mathrm{P}(X_i = 0)\mathrm{E}(Y_i^2|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(Y_i^2|X_i = 1)\\ &= (1 - p_x) + p_x\mathrm{E}(Y_i^2|X_i = 1).\end{aligned}$$

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

$$\begin{aligned}\mathrm{E}(Y_i^2) &= (1 - p_x) + p_x\mathrm{E}[(Z + \delta I)^2]\\ &= (1 - p_x) + p_x[\mathrm{E}(Z^2) + 2\mathrm{E}(ZI)\delta + \mathrm{E}(I^2)\delta^2]\\ &= (1 - p_x) + p_x[1 + 2\phi(\tau)\delta + p_\tau\delta^2]\\ &= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau \delta^2,\end{aligned}$$

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

$$\begin{aligned}\mathrm{Var}(Y_i) &= \mathrm{E}(Y_i^2) - \mathrm{E}(Y_i)^2\\ &= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau \delta^2 - (p_x p_\tau \delta)^2\\ &= 1 + 2p_x\phi(\tau)\delta + p_x p_\tau(1 - p_x p_\tau)\delta^2.\end{aligned}$$

For the binary variable, we first calculate

$$\begin{aligned}\mathrm{E}(T_i) &= \mathrm{P}(X_i = 0)\mathrm{E}(T_i|X_i = 0) + \mathrm{P}(X_i = 1)\mathrm{E}(T_i|X_i = 1)\\ &= (1 - p_x)[1 - \Phi(t)] + p_x[1 - \Phi(t - \delta)]\\ &= 1 - (1 - p_x)\Phi(t) - p_x\Phi(t - \delta).\end{aligned}$$

This implies that

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

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

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

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

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

$$= p_\tau \delta.$$

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

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

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

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.

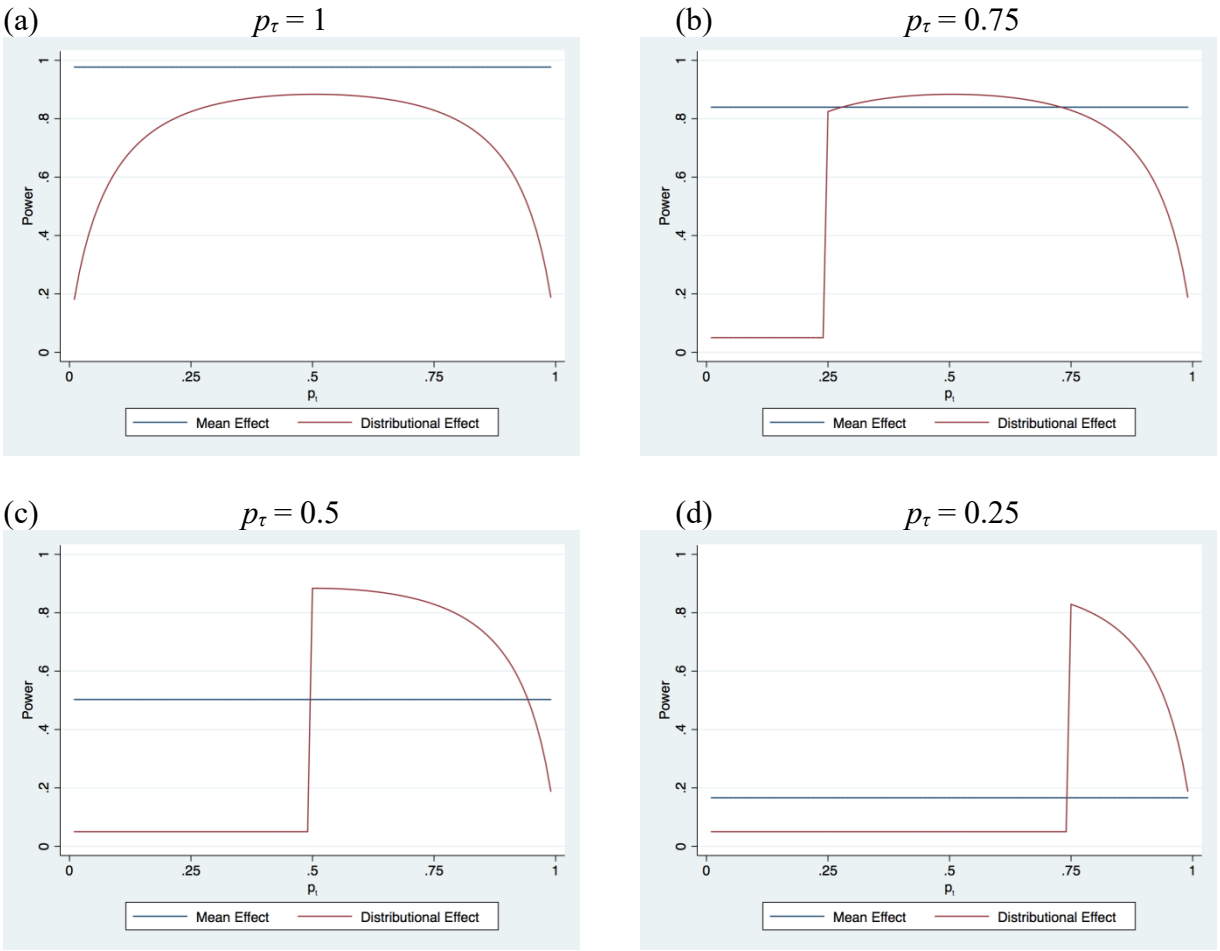
We do not calculate the power of a test for values of t in $[\tau,\ \tau + \delta]$ but rather note that the power will be somewhere between the cases when t is above or below this interval.

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 A1: 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 B

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

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

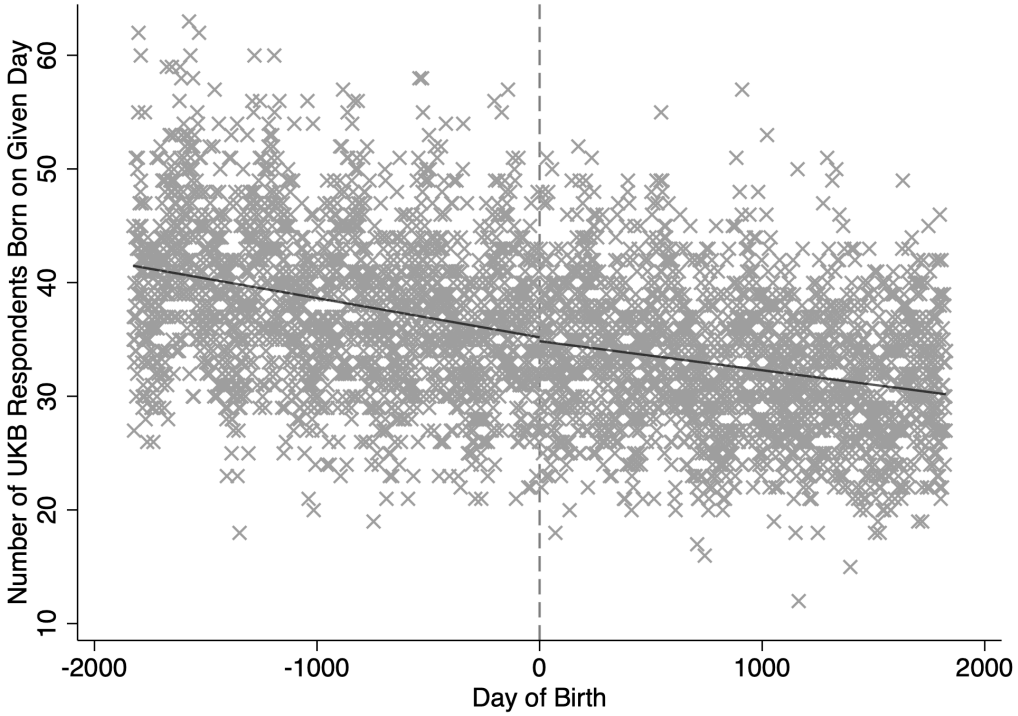
So for small μ, we have

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

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.

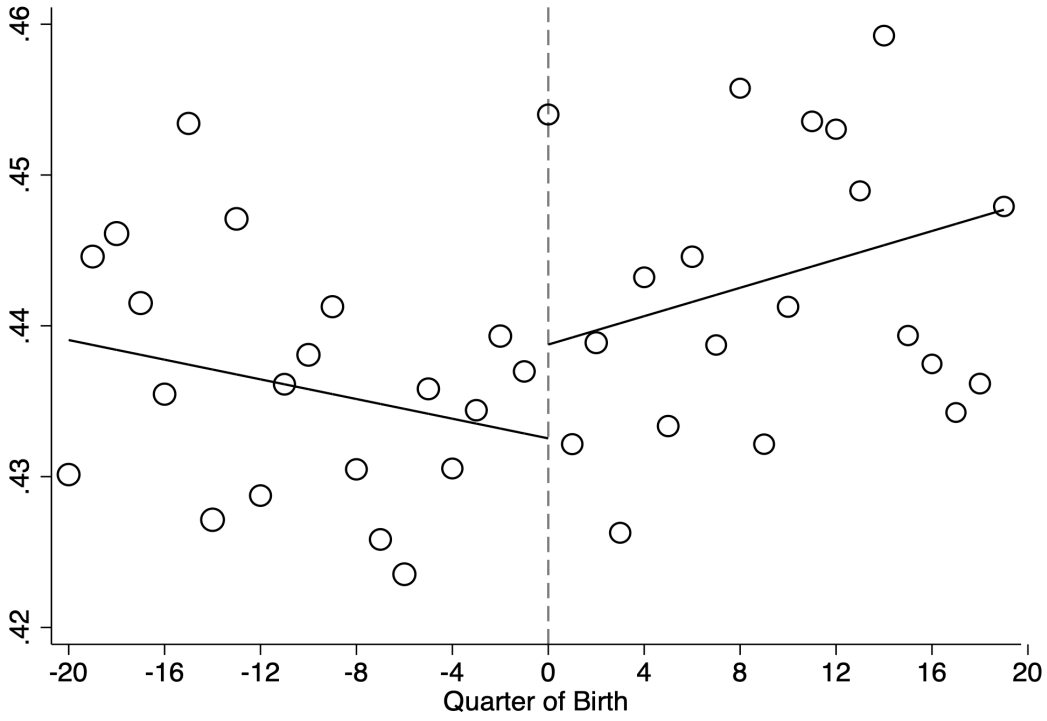
Appendix C

Appendix Figure C1: McCrary Test



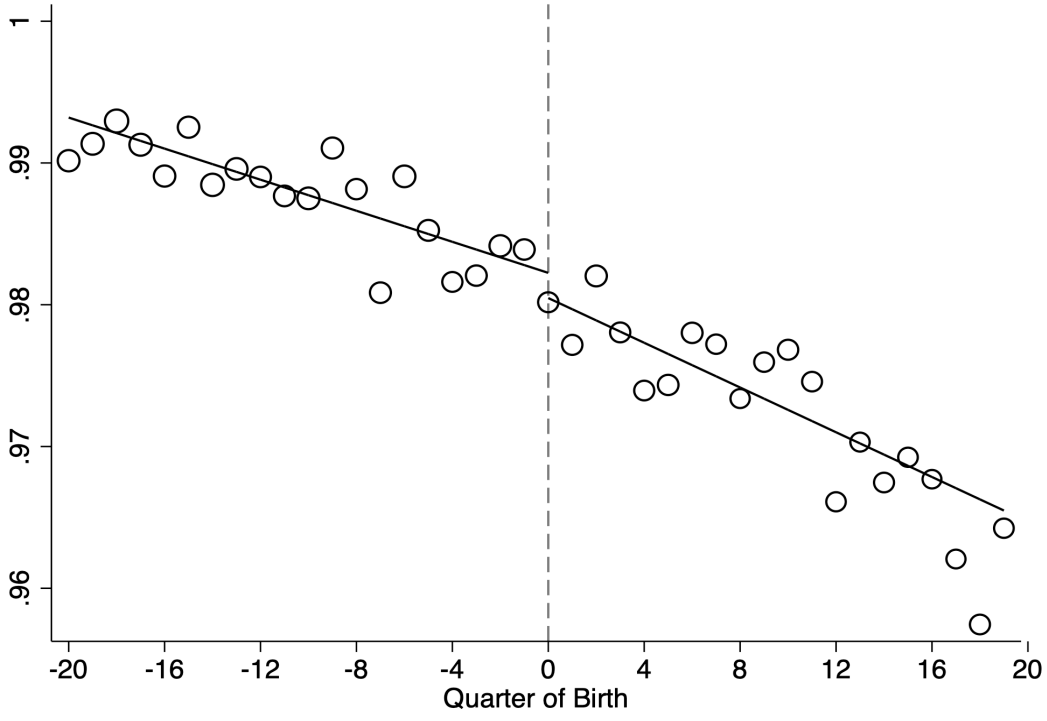
Notes: The figure shows the fraction of study participants by day of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at age 15. The curves show nonparametric birth cohort trends. The estimated discontinuity of the density is -0.0365 with a standard error of 0.0227. $N = 129{,}370$.

Appendix Figure C2: Male



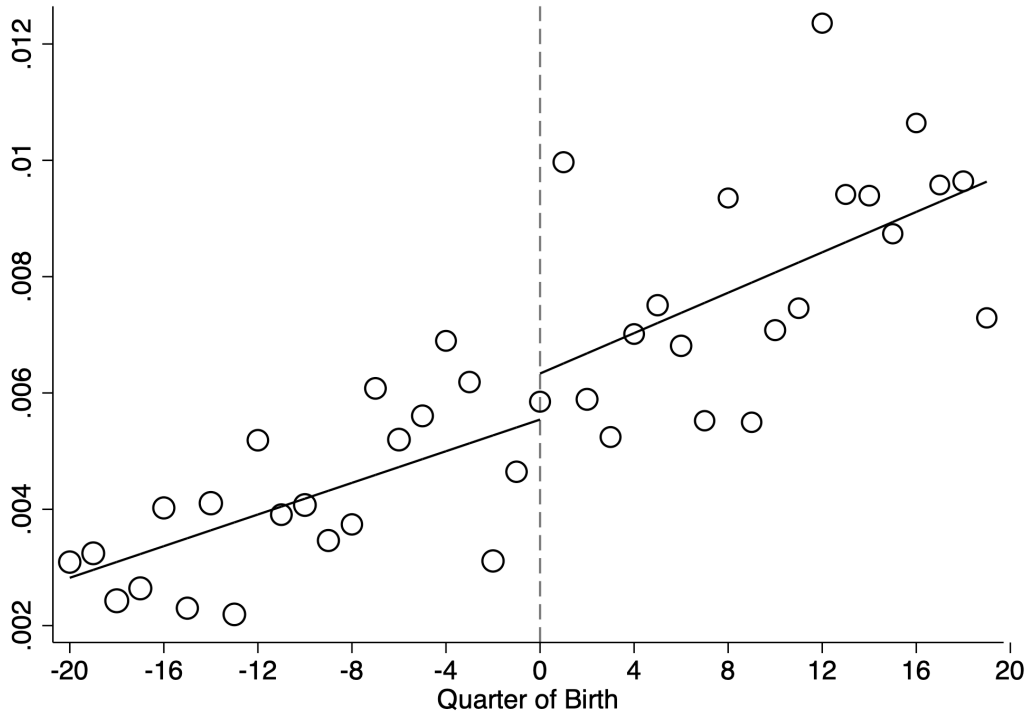
Notes: The figure shows the fraction of male study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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 = 129,370.

Appendix Figure C3: White



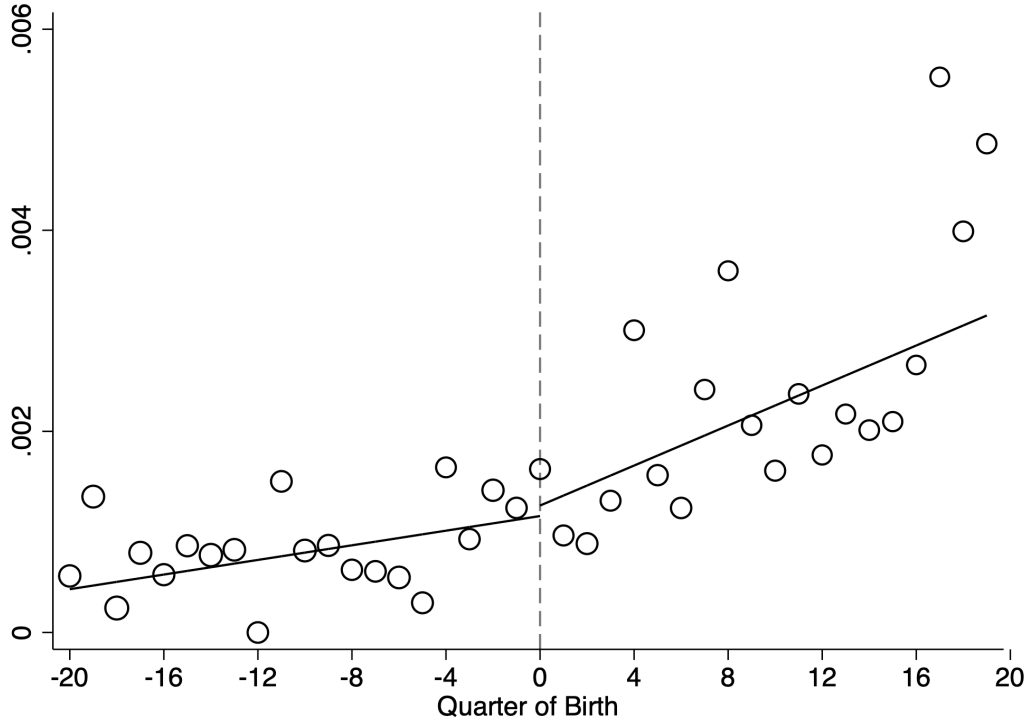
Notes: The figure shows the fraction of white study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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 = 129,370.

Appendix Figure C4: Mixed Ethnicity



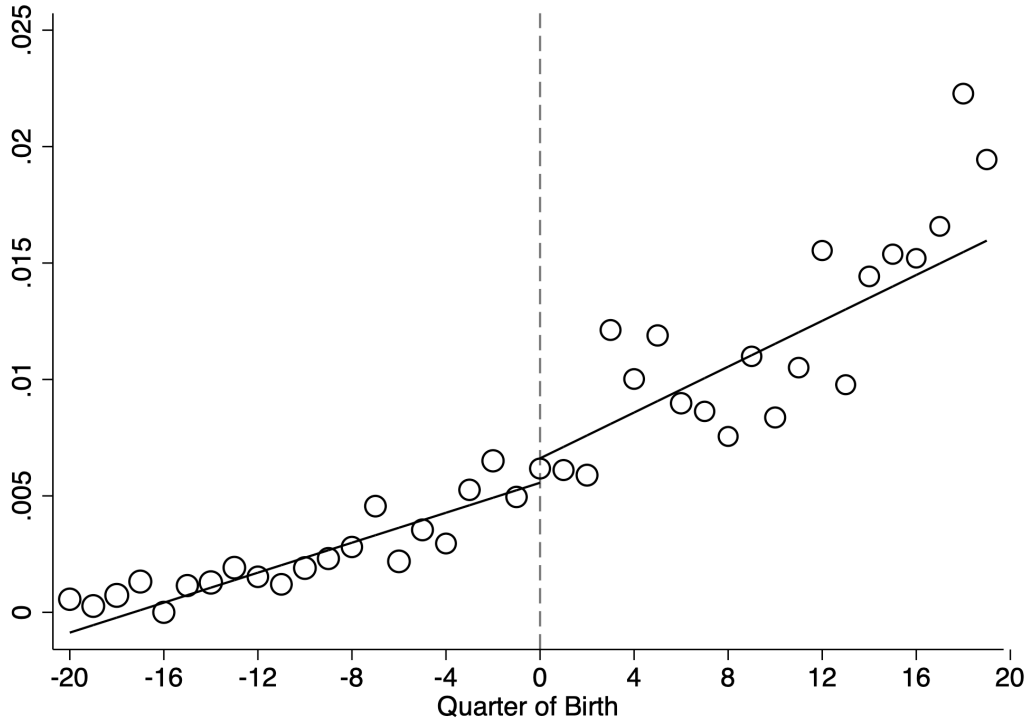
Notes: The figure shows the fraction of study participants of mixed ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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 = 129,370.

Appendix Figure C5: Asian



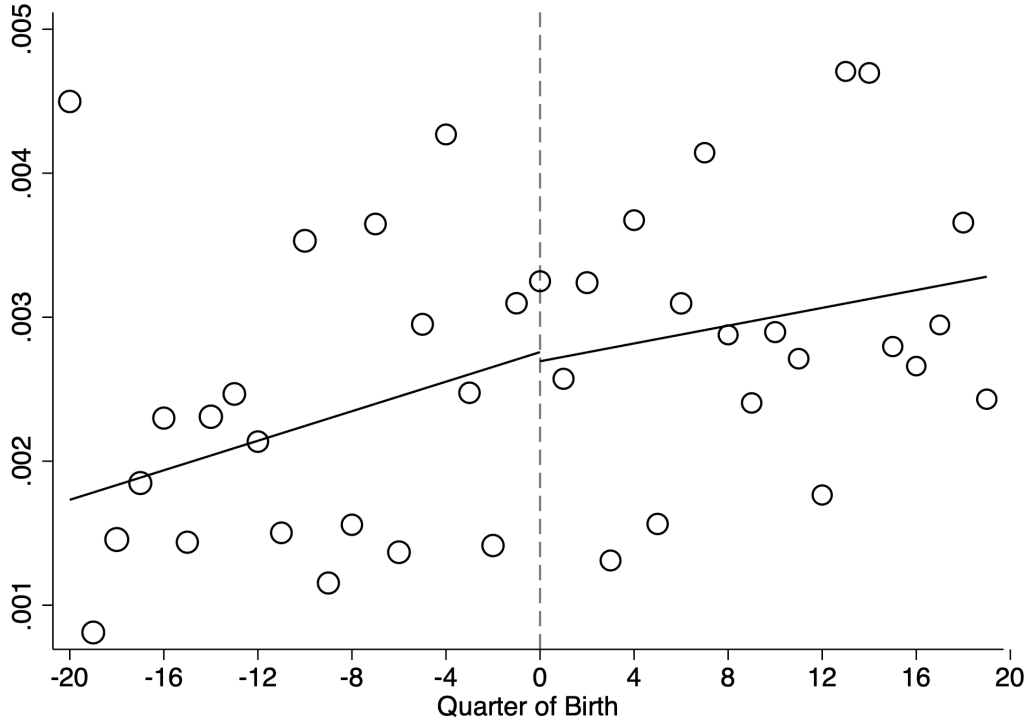
Notes: The figure shows the fraction of Asian study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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 = 129,370.

Appendix Figure C6: Black



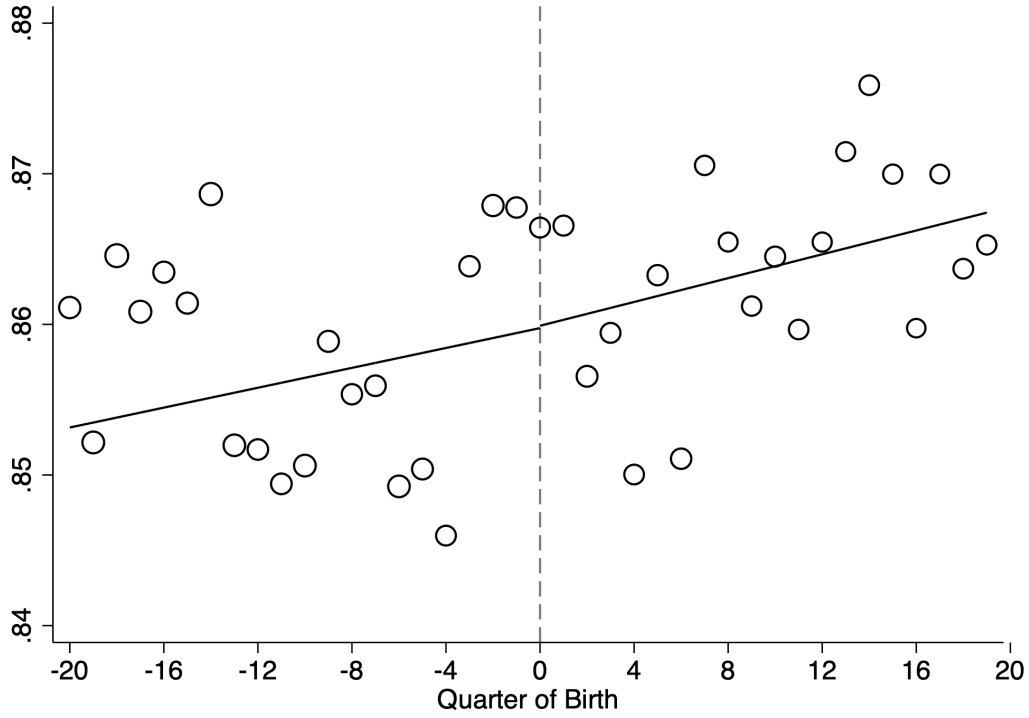
Notes: The figure shows the fraction of black study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave at 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 = 129,370.

Appendix Figure C7: Other Ethnicity



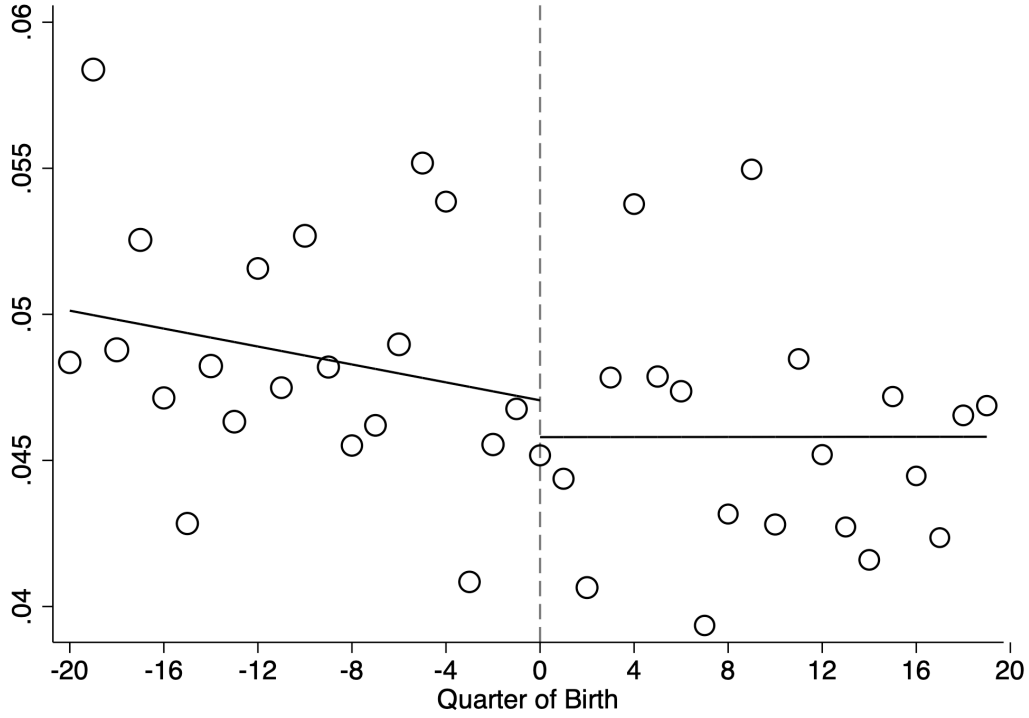
Notes: The figure shows the fraction of study participants of another ethnicity by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before 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 = 129,370.

Appendix Figure C8: Born in England



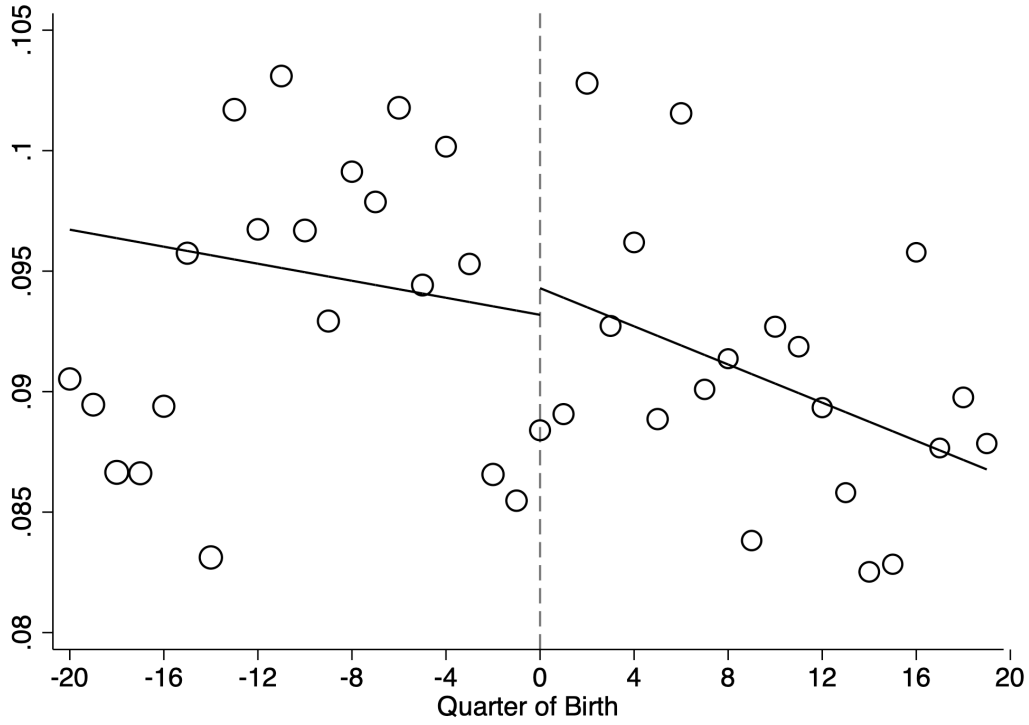
Notes: The figure shows the fraction of study participants born in England by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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 = 129,370.

Appendix Figure C9: Born in Wales



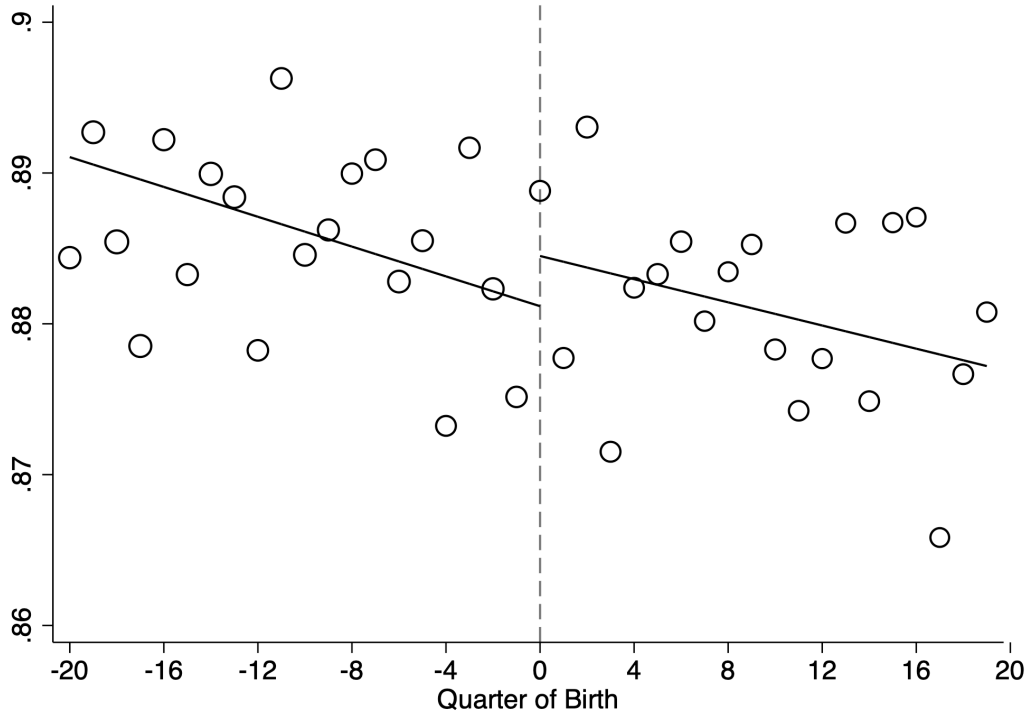
Notes: The figure shows the fraction of study participants born in Wales by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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 = 129,370.

Appendix Figure C10: Born in Scotland



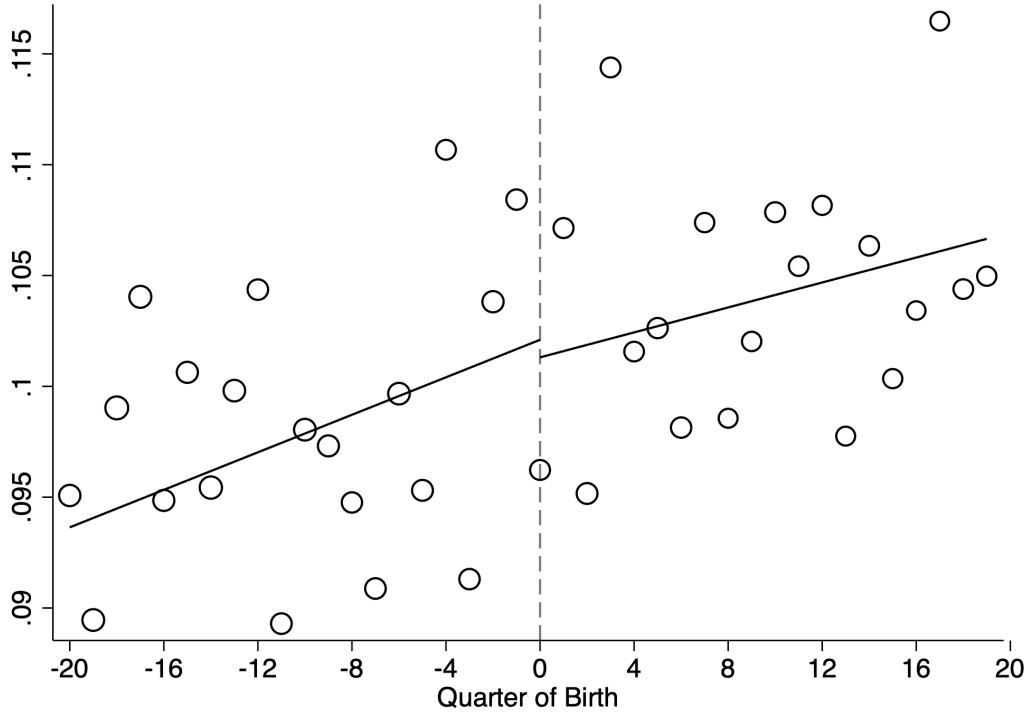
Notes: The figure shows the fraction of study participants born in Scotland by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could 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 = 129,370.

Appendix Figure C11: Right Handed



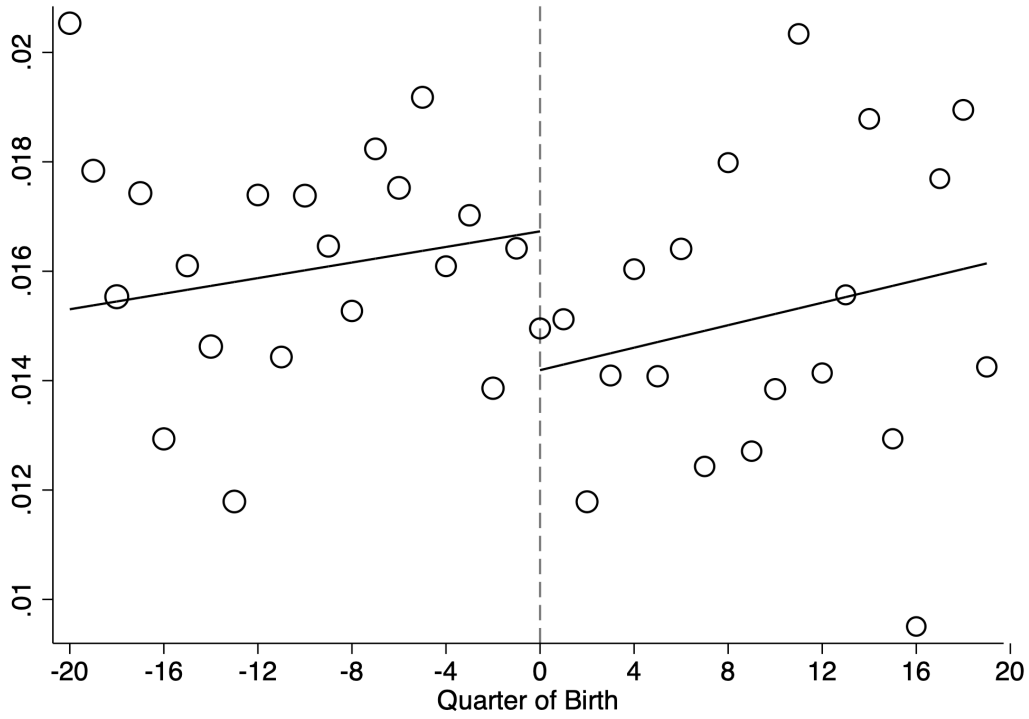
Notes: The figure shows the fraction of right-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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 = 129,341.

Appendix Figure C12: Left Handed



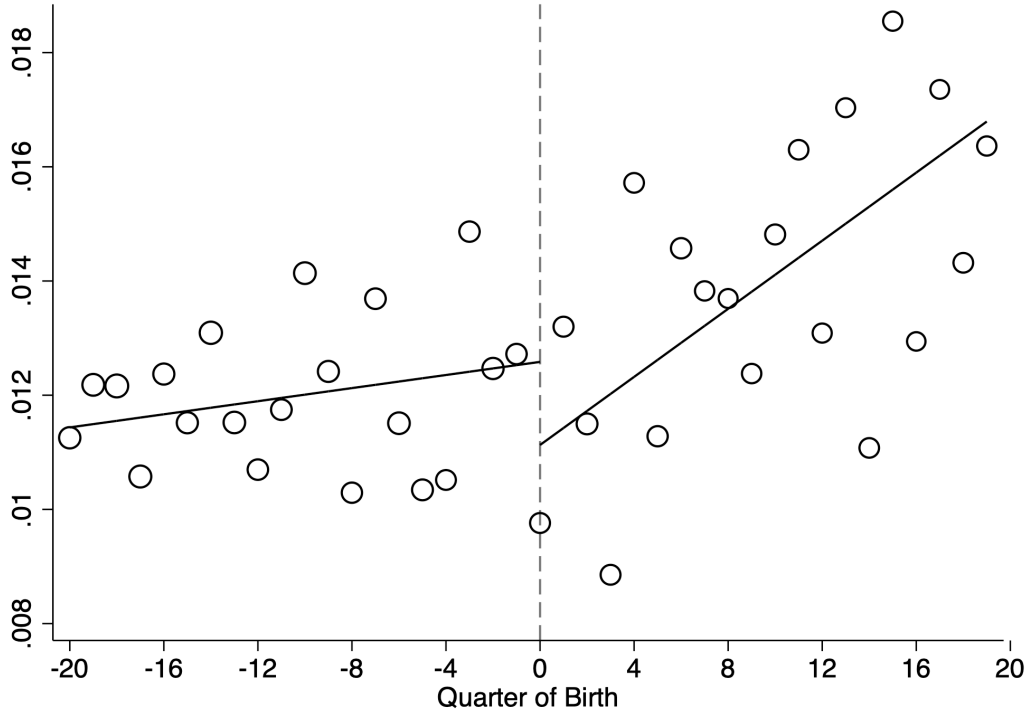
Notes: The figure shows the fraction of left-handed study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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 = 129,341.

Appendix Figure C13: Ambidextrous



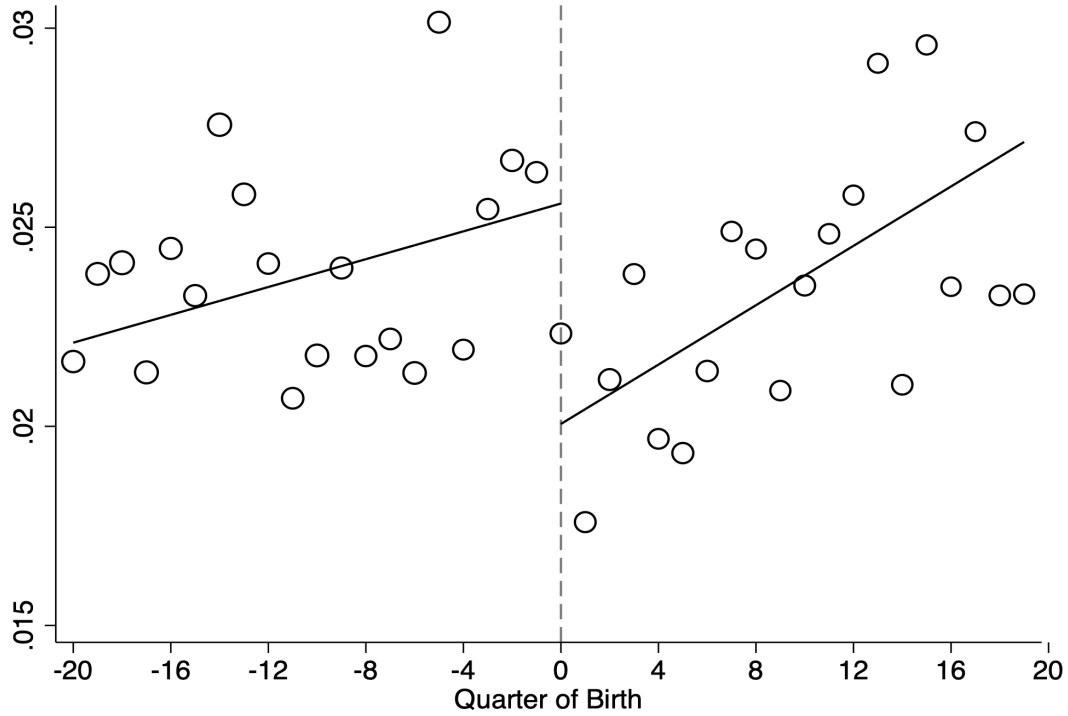
Notes: The figure shows the fraction of ambidextrous study participants by quarter of birth. The dashed vertical line marks the first birth cohort affected by the 1972 school-leaving age reform. Cohorts born to the right of the line had to stay in school until age 16 while cohorts born before could leave 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 = 129,341.

Appendix Figure C14: 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 = 129,197.

Appendix Figure C15: 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 = 127,591.

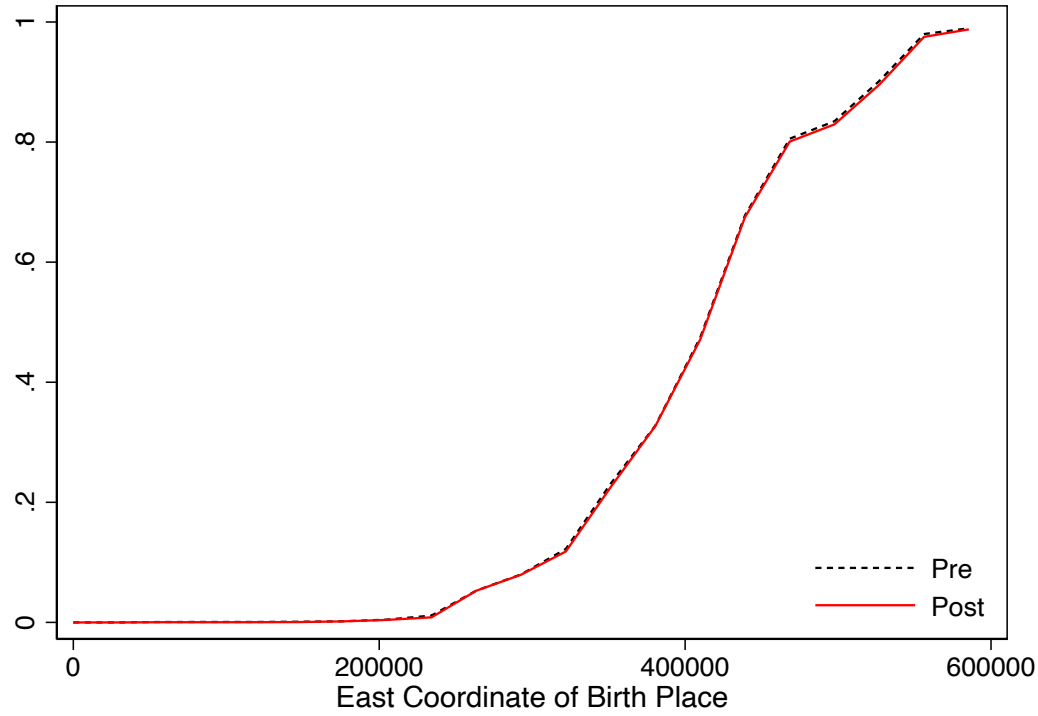
Appendix Table C1: Balance Control Test

	Male	*White*	*Mixed Ethnicity*	*Asian*	*Black*	*Other Ethnicity*	*Born in England*
Post	0.006	-0.002	0.001	0.000	0.001	-0.000	0.000
	[0.006]	[0.002]	[0.001]	[0.000]	[0.001]	[0.001]	[0.004]
N	129,370	129,370	129,370	129,370	129,370	129,370	129,370
Mean of Y	0.436	0.983	0.00507	0.00131	0.00504	0.00273	0.862

	Born in Wales	*Born in Scotland*	*Right Handed*	*Left Handed*	*Ambi-dextrous*	*Adopted*	*Twin*
Post	-0.001	0.001	0.003	-0.001	-0.002	-0.002	-0.006
	[0.003]	[0.004]	[0.004]	[0.004]	[0.001]	[0.001]	[0.002]***
N	129,370	129,370	129,341	129,341	129,341	129,197	127,591
Mean of Y	0.0465	0.0913	0.881	0.103	0.0158	0.0127	0.0253

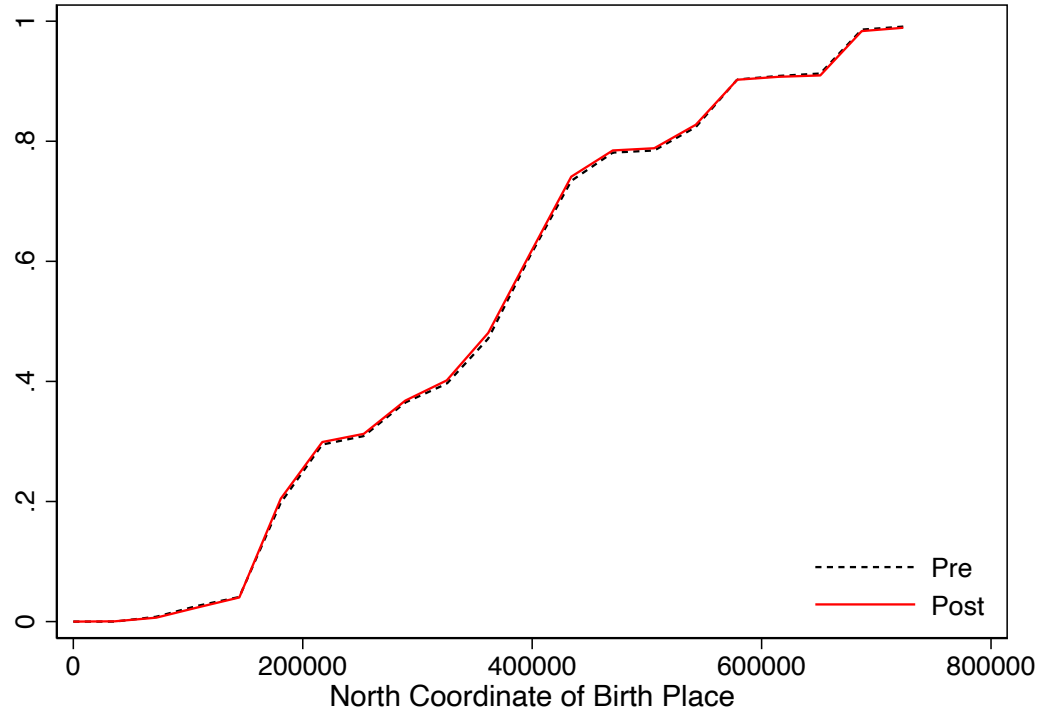
Notes: The table investigates whether predetermined characteristics are smooth are around the September 1, 1957 cutoff. It reports the coefficient on an indicator for being born on or after September 1, 1957 (i.e., "Post") from regressions where the dependent variables is listed in the column. The regressions 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 C16: East Coordinate of Birth Place



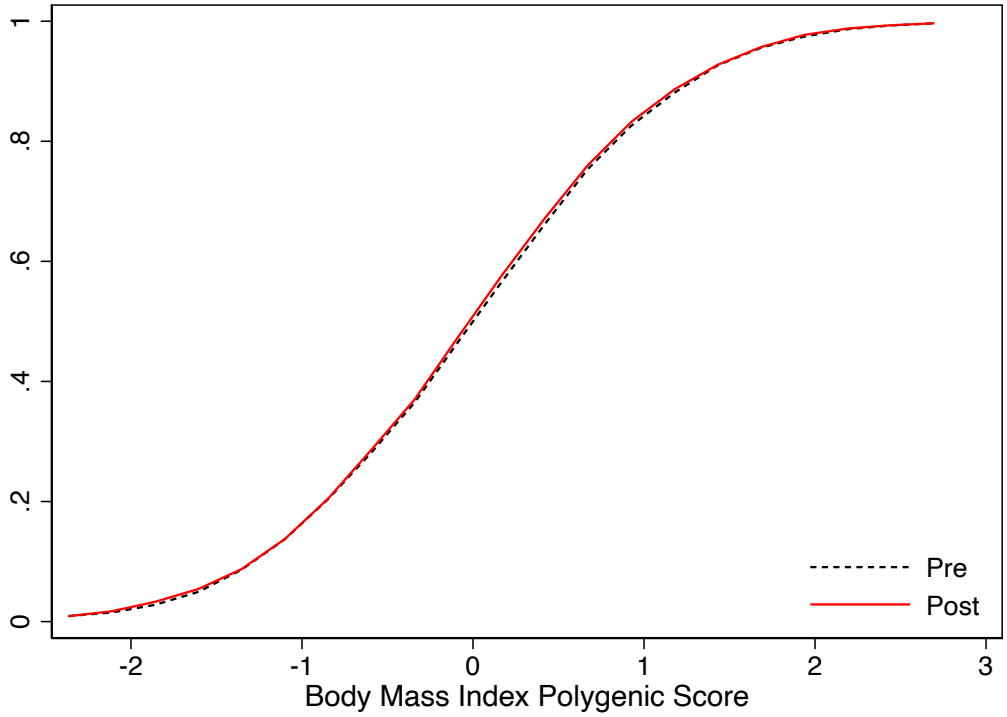
Notes: The figure shows the pre- and post-reform CDFs of east coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 119{,}671$.

Appendix Figure C17: North Coordinate of Birth Place



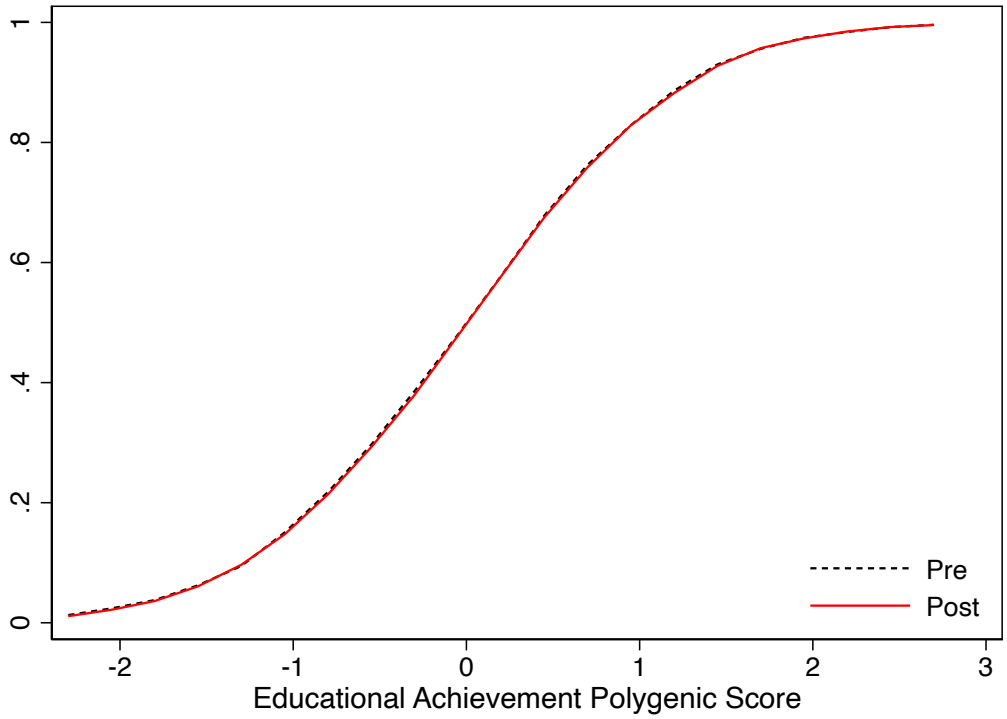
Notes: The figure shows the pre- and post-reform CDFs of north coordinate of place of birth. The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. The *post-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the right. $N = 119{,}671$.

Appendix Figure C18: 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. $N = 119{,}671$.

Appendix Figure C19: 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. $N = 119{,}671$.

Appendix Table C2: Distributional Test

Coordinates of Birth Place		Polygenic Scores	
East	*North*	*BMI*	*Educational Achievement*
0.45	0.45	0.21	0.82

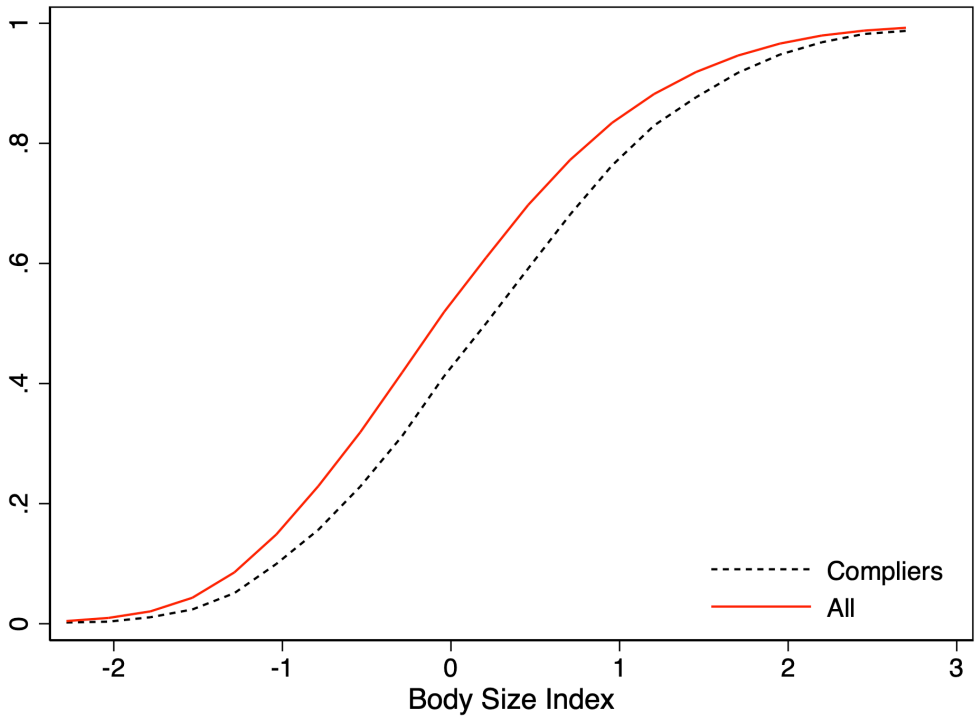
Notes: The table show the p-values of tests of the equality of the pre- and post-reform CDFs. N = 119,671

Appendix D

Appendix Figure D1: Map with Locations of 22 Assessment Centers

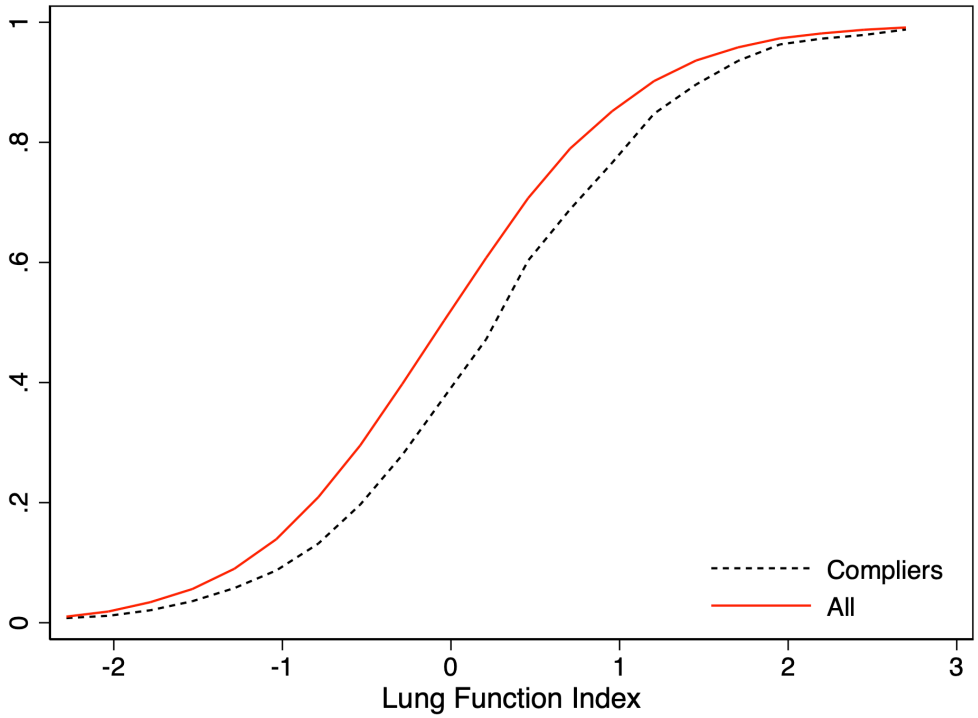
Notes: The figure shows the location of the 22 assessment centers (as well as the location of the pilot study).

Appendix Figure D2: Pre-Reform Cumulative Distribution of Body Size Index For ROSLA Compliers and for Entire UK Biobank Sample



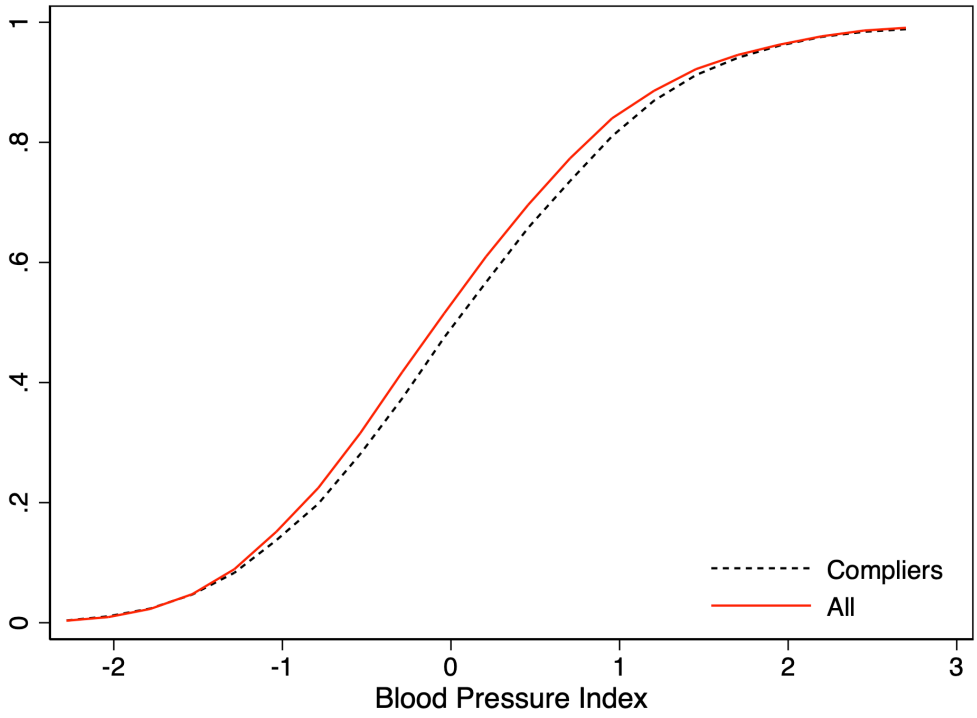
Notes: The figure shows the pre-reform CDFs of body size index for compliers (black dashed) and for the entire population (red solid). The *pre-reform CDF* is the CDF in the limit when date of birth is converging to September 1, 1957 from the left. $N = 11{,}697$ (compliers) and 68,868 (all).

Appendix Figure D3: Pre-Reform Cumulative Distribution of Lung Function Index for ROSLA Compliers and for Entire UK Biobank Sample



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 = 8,848 (compliers) and 55,861 (all).

Appendix Figure D4: Pre-Reform Cumulative Distribution of Blood Pressure Index for ROSLA Compliers and for Entire UK Biobank Sample



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 = 11,919 (compliers) and 69,896 (all).

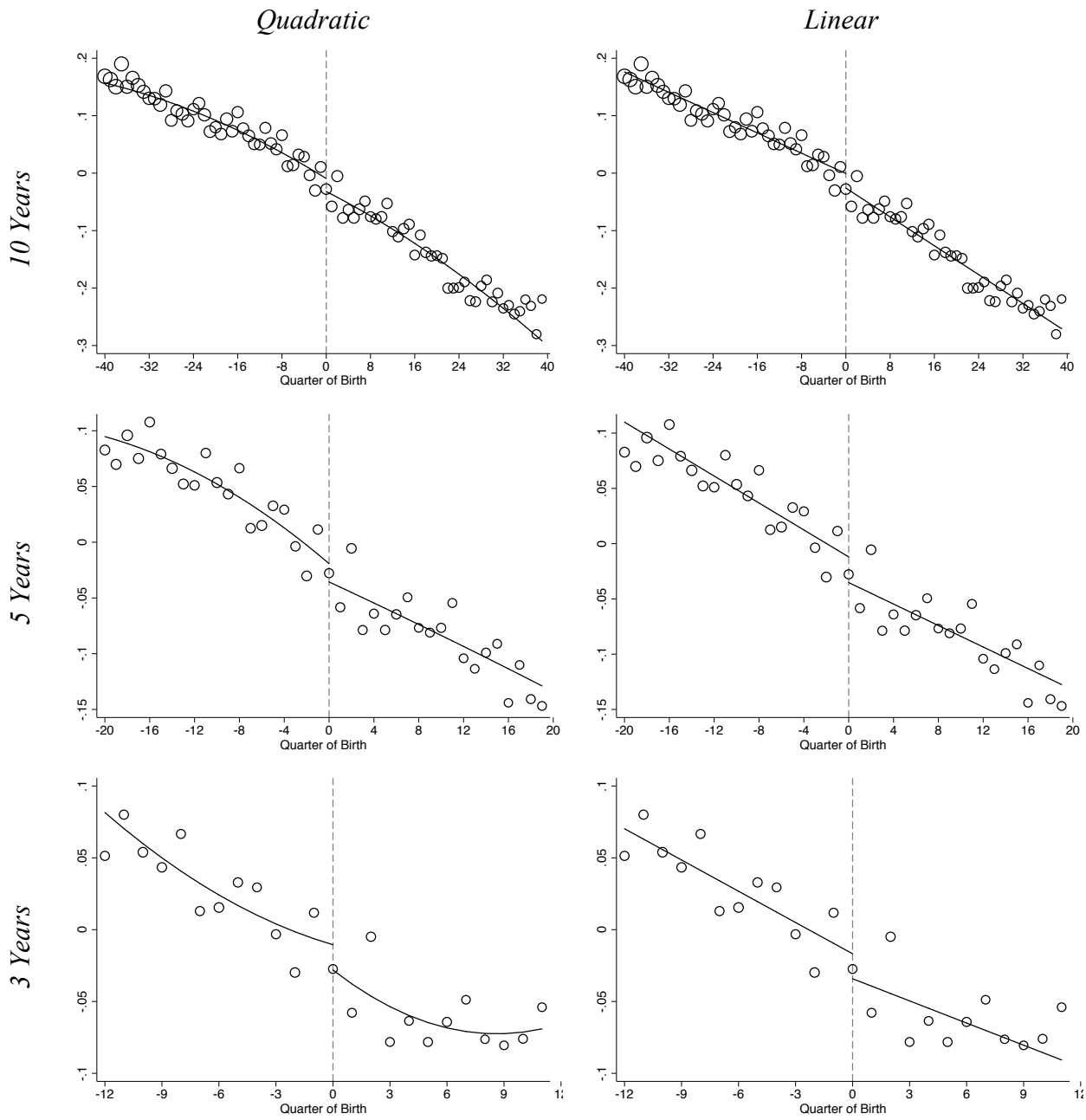
Appendix Table D1: Comparison with Clark and Royer (2010) and Davies et al. (2018)

	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	1.66	0.01	2.42	-0.49	0.01	-0.08
Lower Bound 95% CI	-0.16	-0.06	-0.58	-1.37	-0.07	-0.16
Upper Bound 95% CI	3.49	0.09	5.42	0.40	0.09	0.00
N	129,165	129,165	129,165	128,856	128,856	128,856
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.

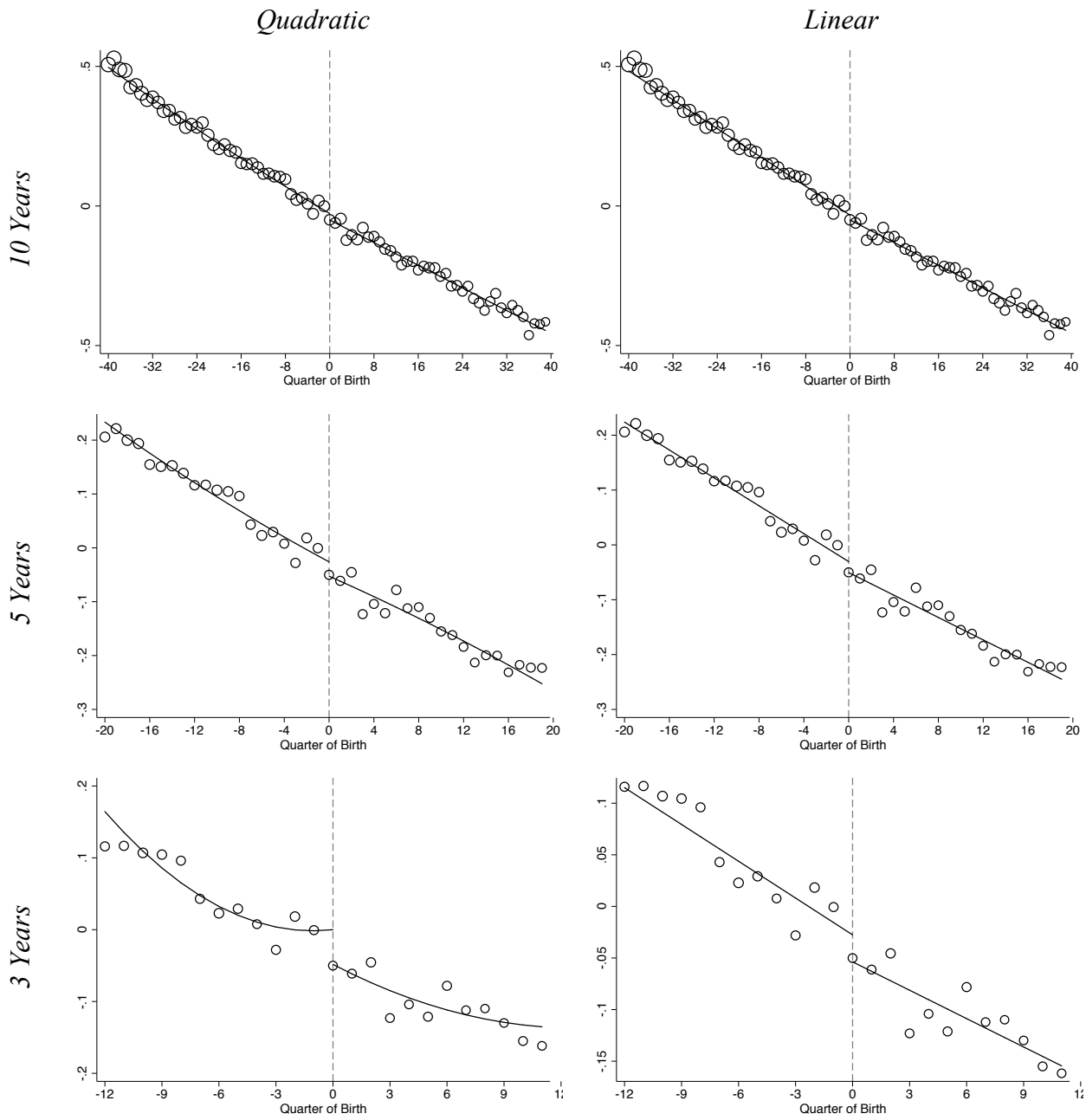
Appendix E

Appendix Figure E1: Average of Body Size Index by Quarter of Birth



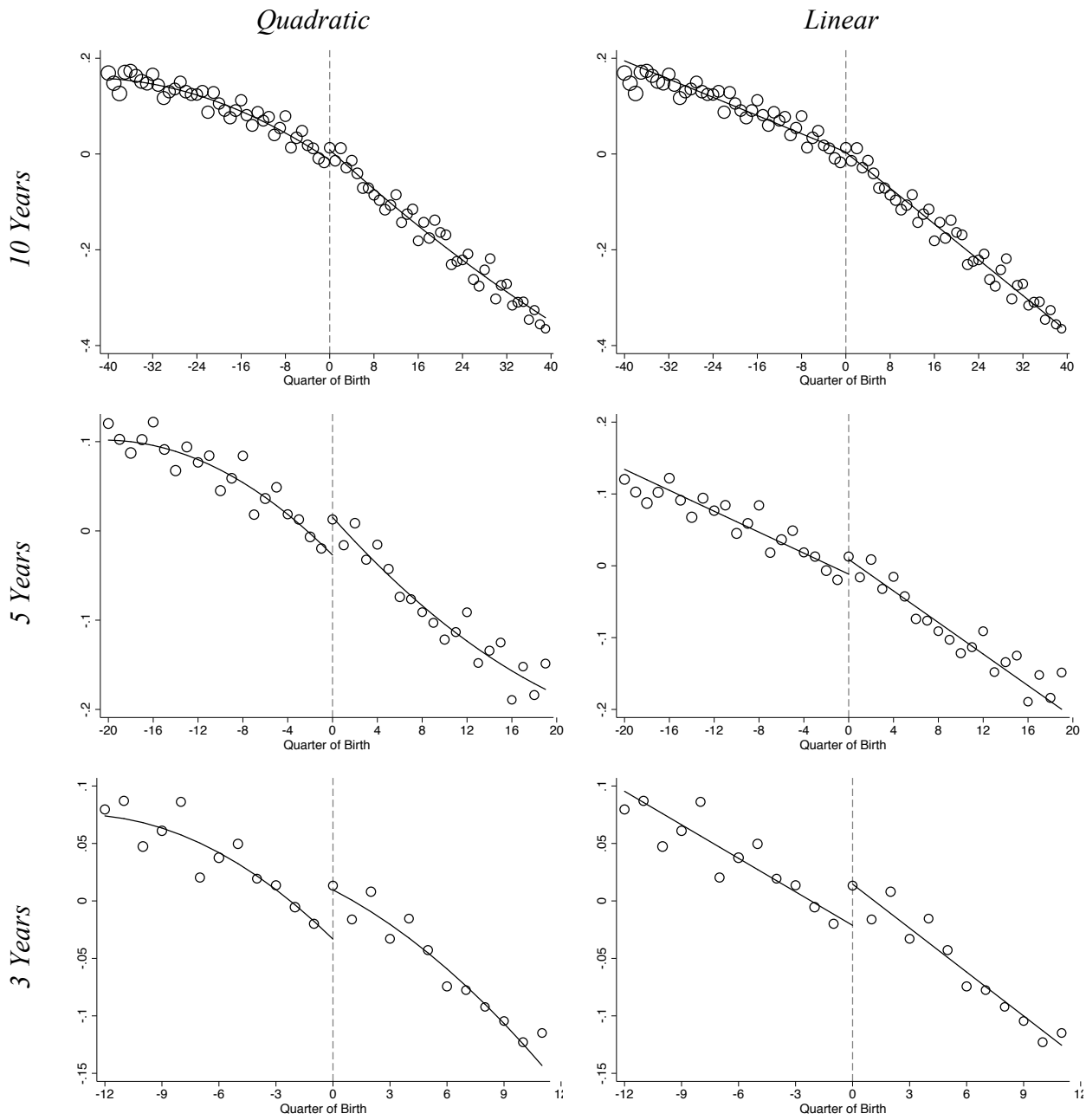
Notes: The figure assesses the sensitivity of the results for body size index to the choice of bandwidth and to the use of linear trends. It shows the average of body size index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column uses linear 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 E2: Average of Lung Function Index by Quarter of Birth



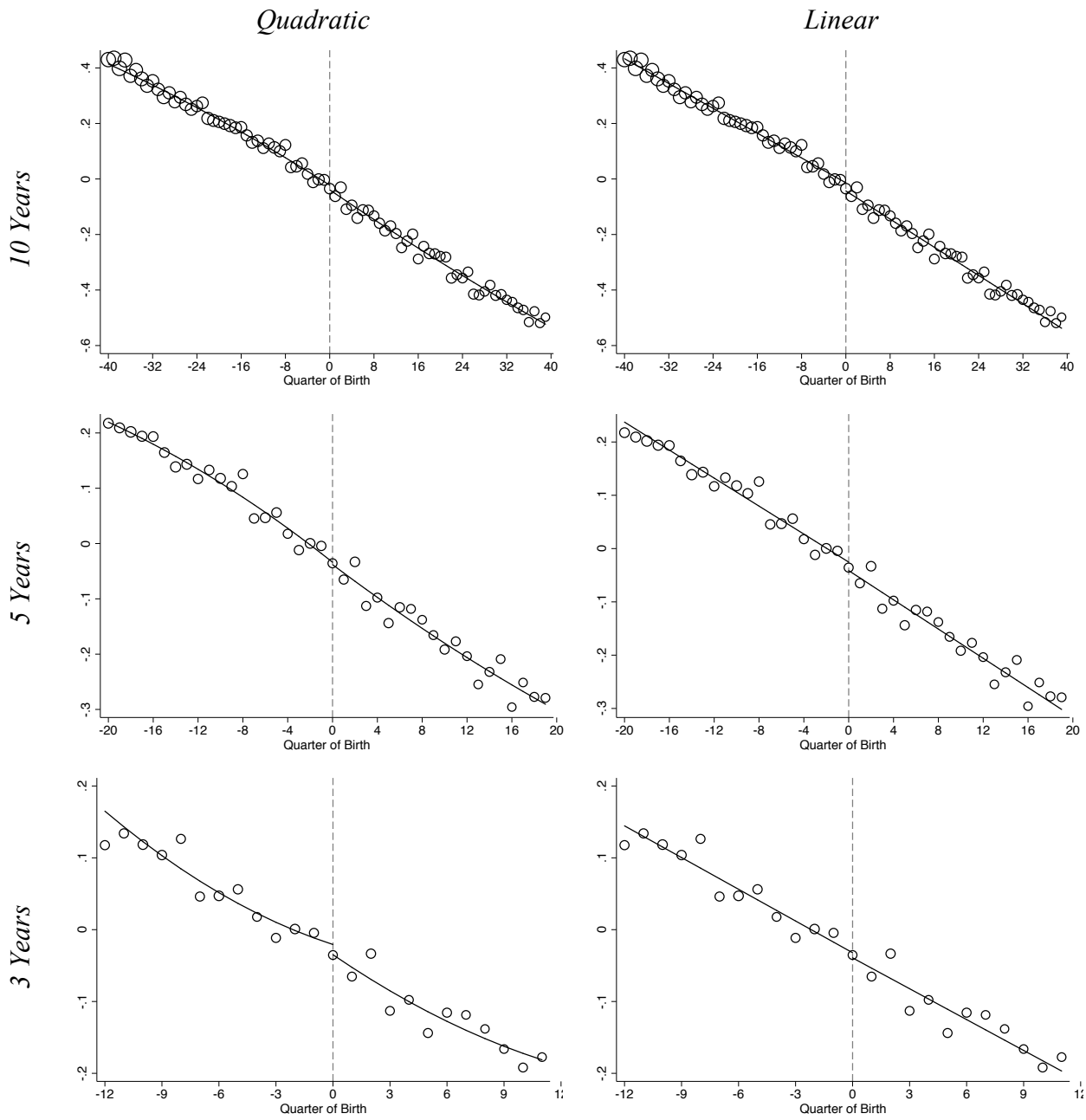
Notes: The figure assesses the sensititivity of the results for lung function index to the choice of bandwidth and to the use of linear trends. It shows the average of lung function index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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 E3: Average of Blood Pressure Index by Quarter of Birth



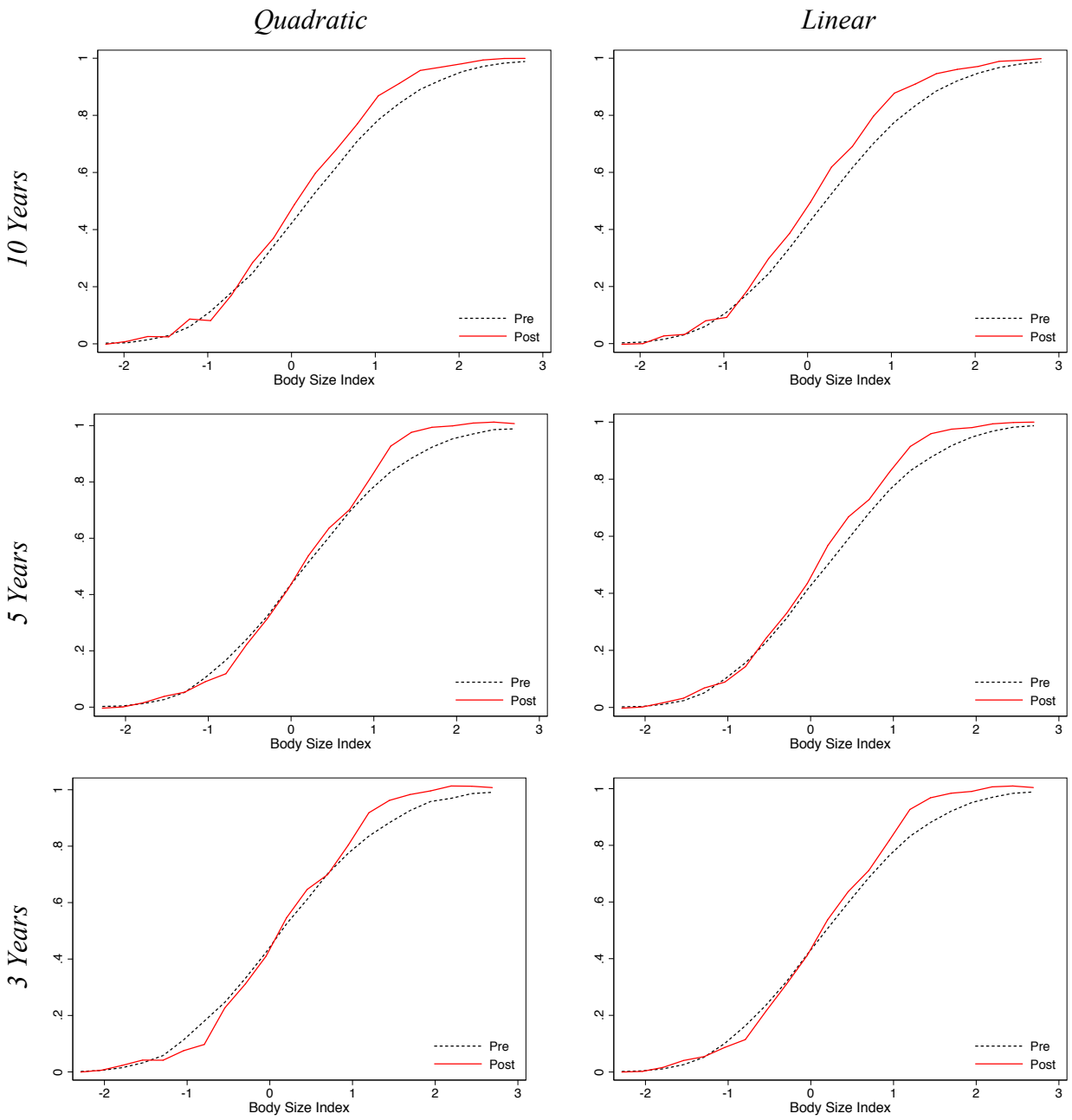
Notes: The figure assesses the sensitivity of the results for blood pressure index to the choice of bandwidth and to the use of linear trends. It shows the average of blood pressure index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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 E4: Average of Summary Index by Quarter of Birth



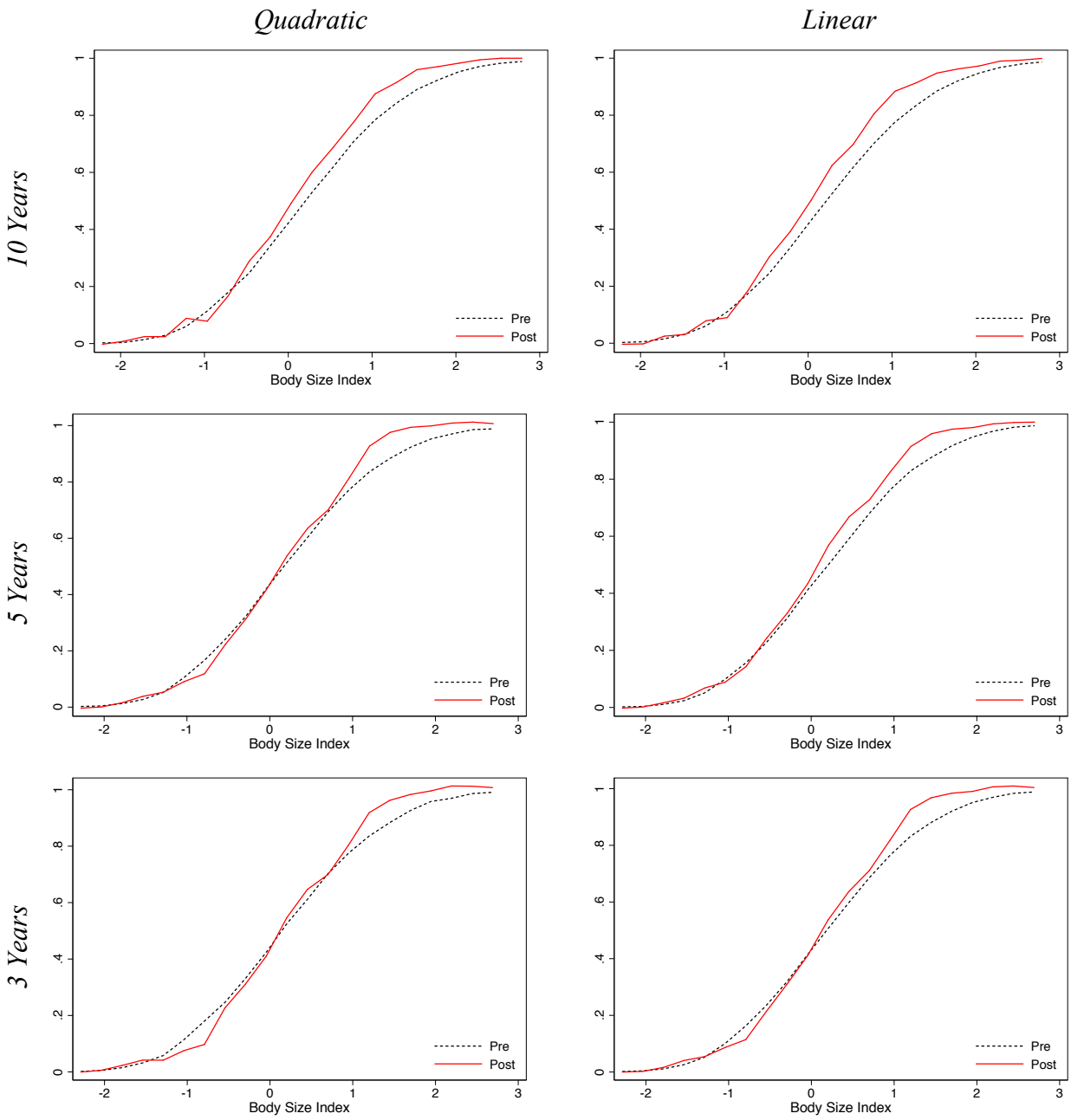
Notes: The figure assesses the sensititivity of the results for the summary index to the choice of bandwidth and to the use of linear trends. It shows the average of the summary index by quarter of birth. The left-hand side column uses quadratic trends in quarter of birth. The right-hand side column 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 E5: Distributional Effects on Body Size (No Controls)



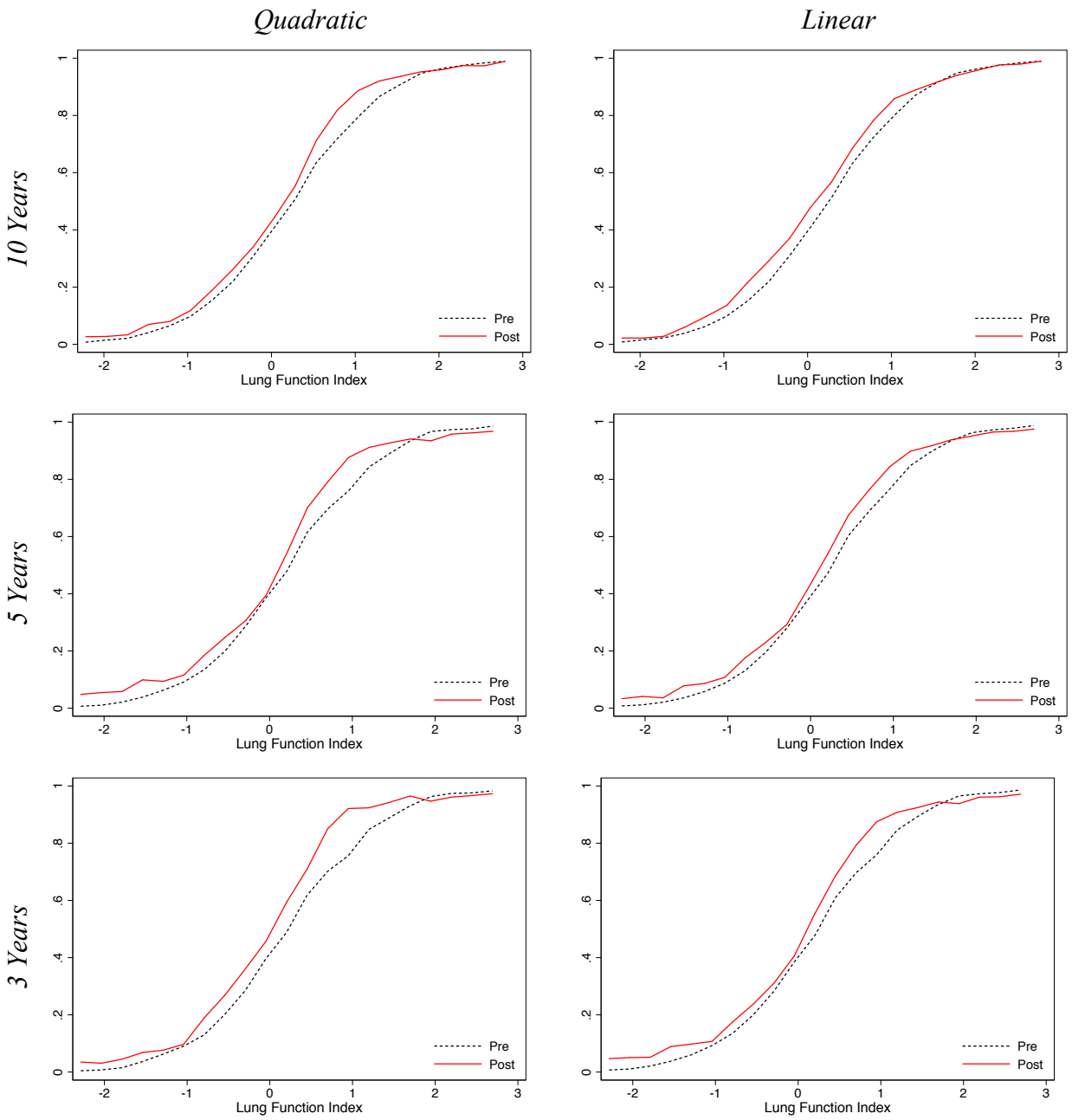
Notes: The figure assesses the sensitivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The 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 E6: Distributional Effects on Body Size (With Controls)



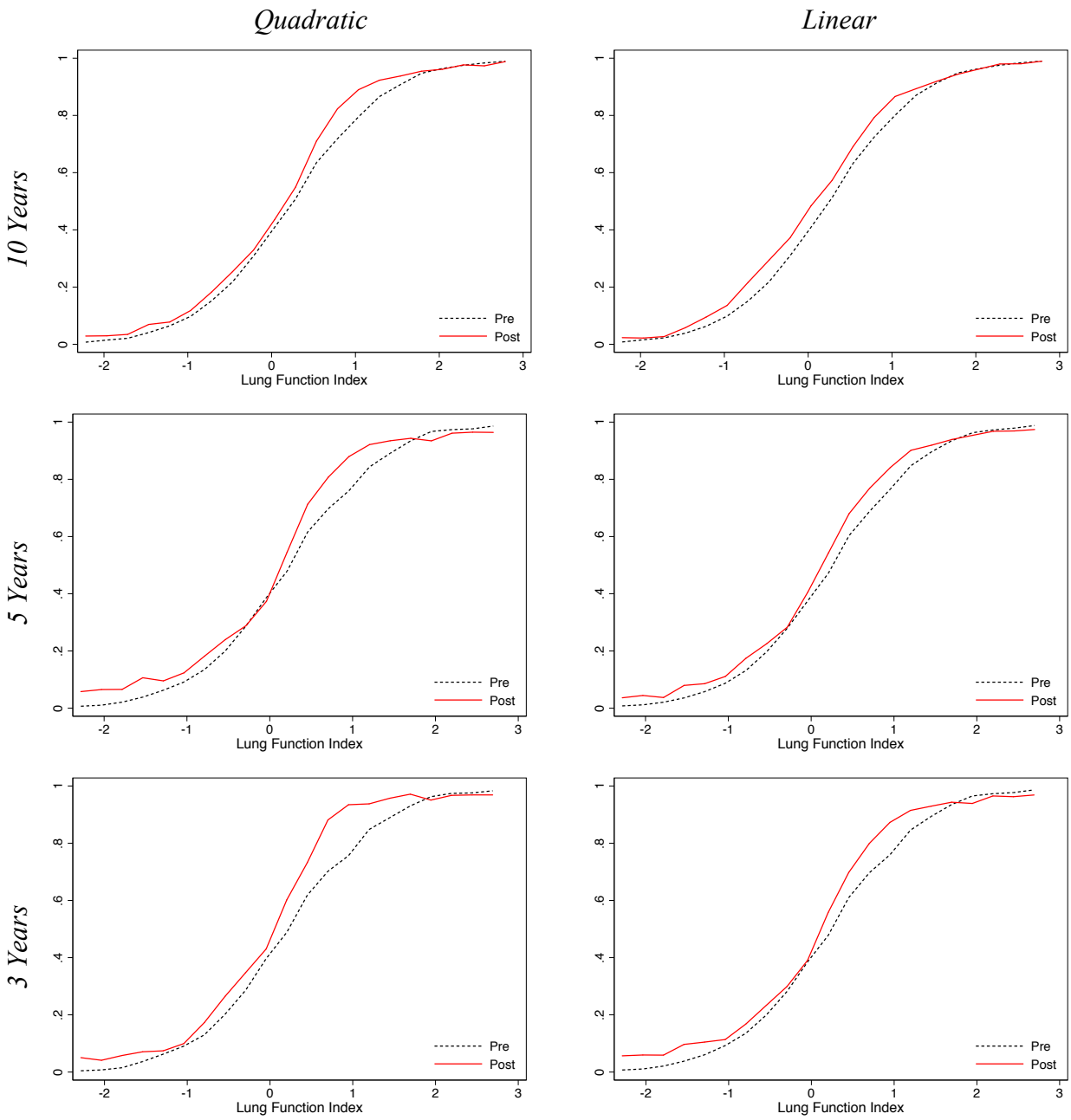
Notes: The figure assesses the sensitivity of the distributional effects on the body size index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the body size index. The left-hand side column uses quadratic trends in date of birth. The 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 E7: Distributional Effects on Lung Function (No Controls)



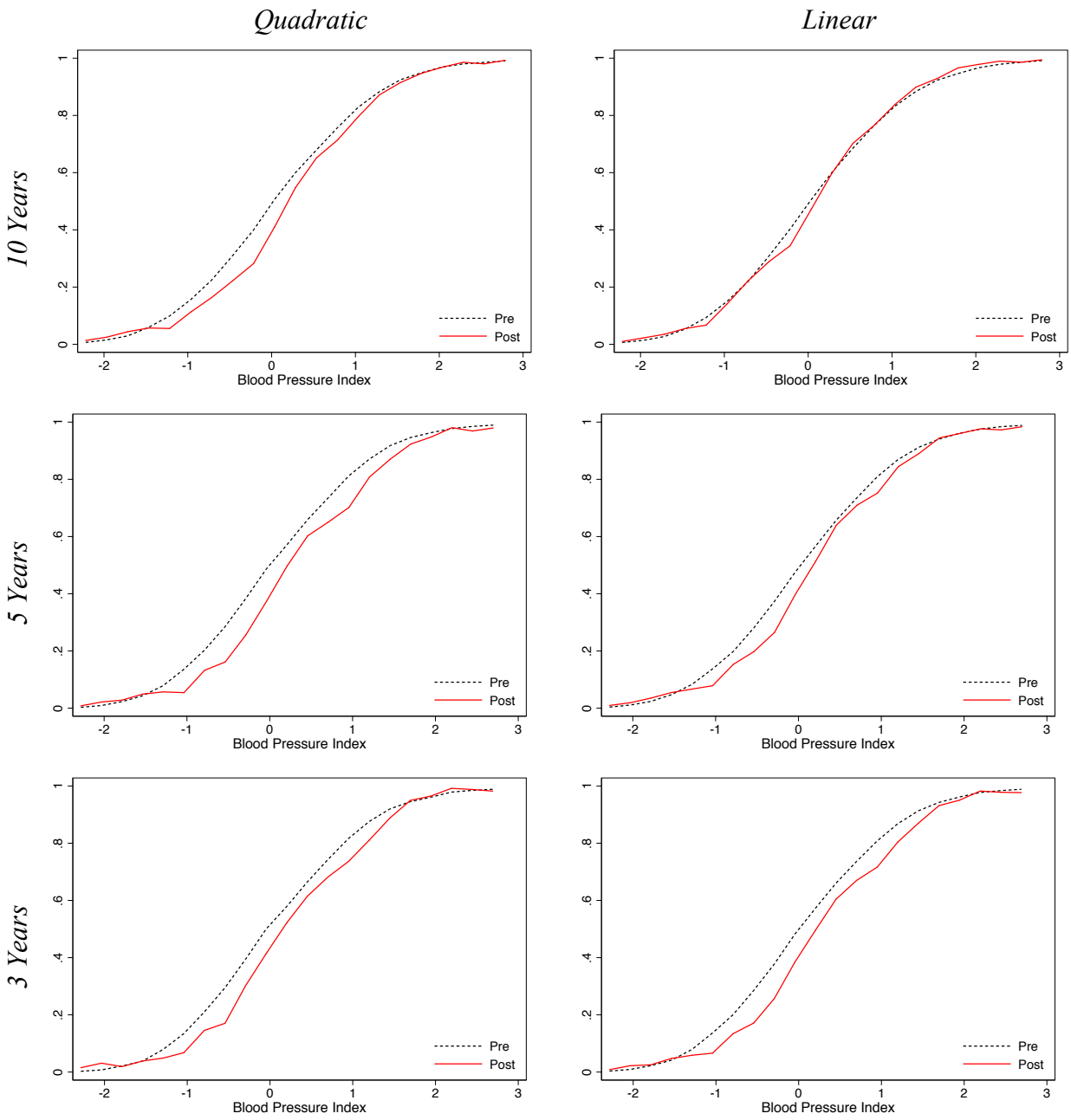
Notes: The figure assesses the sensitivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for compliers of the lung function index. The left-hand side column uses quadratic trends in date of birth. 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 E8: Distributional Effects on Lung Function (With Controls)



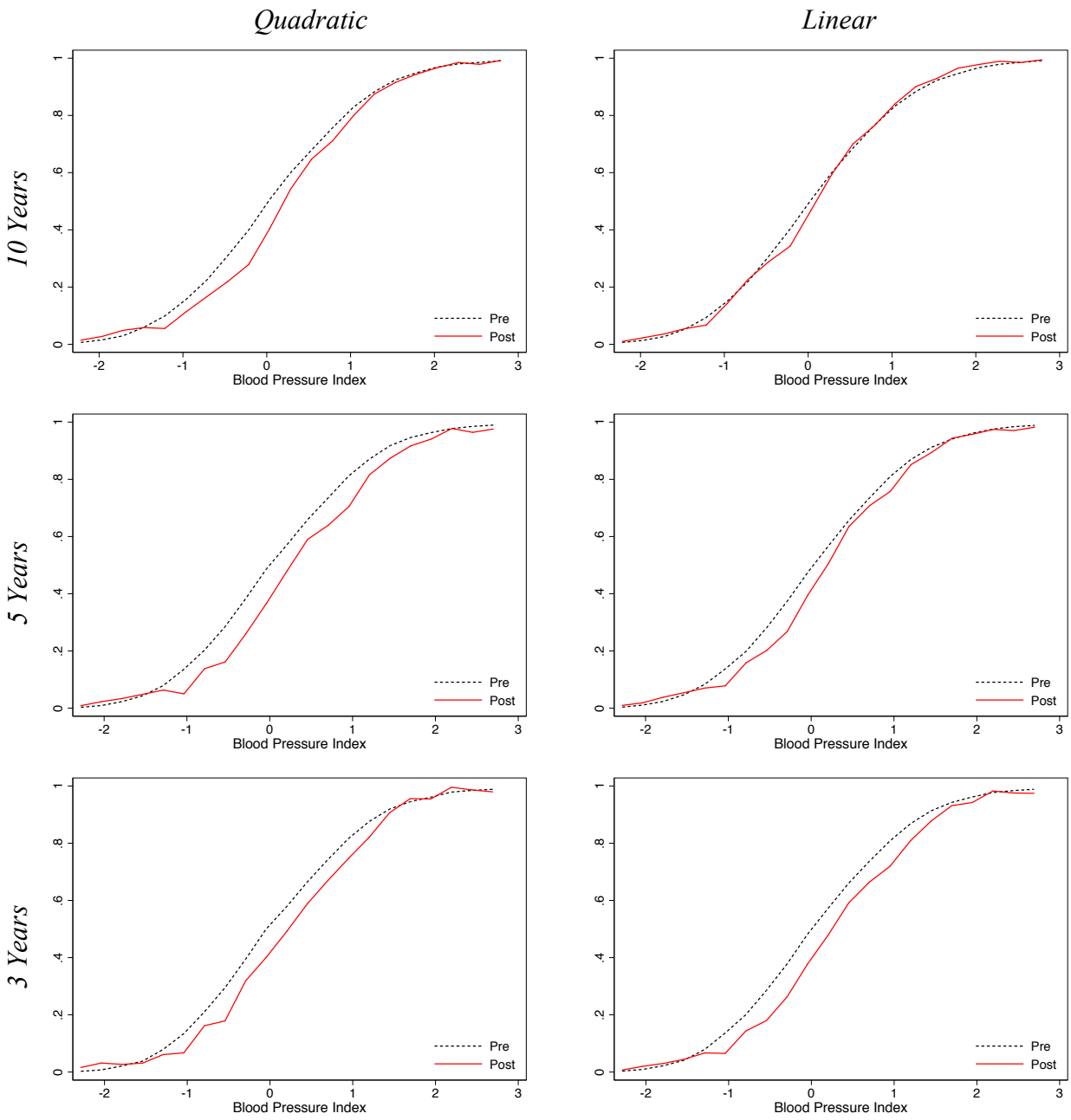
Notes: The figure assesses the sensititivity of the distributional effects on the lung function index to the choice of bandwidth and to the use of linear trends. It shows the pre- and post-reform CDFs for complies of the lung function index. The left-hand side column uses quadratic trends in date of 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 E9: Distributional Effects on Blood Pressure (No Controls)



Notes: The figure assesses the sensitivity 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 compliers 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 E10: Distributional Effects on Blood Pressure (With Controls)



Notes: The figure assesses the sensitivity 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 compliers 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 E1: 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	0.0662	0.3070	0.0060	0.0896
With Controls	0.1656	0.3780	0.0671	0.2598	0.0050	0.0932
Bottom Half						
No Controls	0.9660	0.8154	0.8812	0.9730	0.4158	0.9526
With Controls	0.8186	0.4496	0.9016	0.8514	0.3812	0.9396
Top Half						
No Controls	0.0554	0.3446	0.0129	0.1058	0.0002	0.0126
With Controls	0.0450	0.3266	0.0129	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 E2: 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	0.1989	0.1964	0.0768	0.1712
With Controls	0.1578	0.1572	0.2601	0.2340	0.0706	0.2352
Bottom Half						
No Controls	0.3672	0.4248	0.3672	0.3768	0.0552	0.5962
With Controls	0.3744	0.6082	0.4473	0.4406	0.0626	0.7634
Top Half						
No Controls	0.0626	0.0438	0.1331	0.1168	0.1090	0.0618
With Controls	0.0778	0.0524	0.1609	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 E3: 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	0.0402	0.0262	0.5432	0.0362
With Controls	0.0306	0.2006	0.0568	0.0358	0.5856	0.0532
Bottom Half						
No Controls	0.0100	0.0836	0.0165	0.0126	0.4152	0.0102
With Controls	0.0226	0.1806	0.0277	0.0266	0.4658	0.0172
Top Half						
No Controls	0.0420	0.2772	0.2115	0.0500	0.6924	0.1502
With Controls	0.0380	0.2264	0.2140	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 F

Appendix Table F1: Effect on Distribution of Annual Household Income

	Annual household income below			
	£18,000	*£31,000*	*£52,000*	*£100,000*
Stayed in school until 16	-0.03	-0.13	-0.02	0.02
	[0.031]	[0.043]***	[0.046]	[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* = 116,188.

Appendix Table F2:
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	-0.03	-0.06	-0.01	-0.02	0.00	-0.44	-0.02
	[0.022]	[0.041]	[0.034]	[0.019]	[0.026]	[0.255]*	[0.088]
Mean of Y	0.08	0.41	0.80	0.95	0.90	-1.33	0.00
N Observations	128,950				128,643	129,184	117,942

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 F3: Effect on Diet

	Calories	*% Sugars*	*% Fat*	*% Saturated Fat*	*% Carbo-hydrates*
Stayed in school until 16	-125.07	0.02	-0.03	-0.02	0.01
	[156.711]	[0.018]	[0.017]*	[0.009]**	[0.021]
Mean of Y	2,109	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* = 127,954 observations, corresponding to 58,575 study participants.

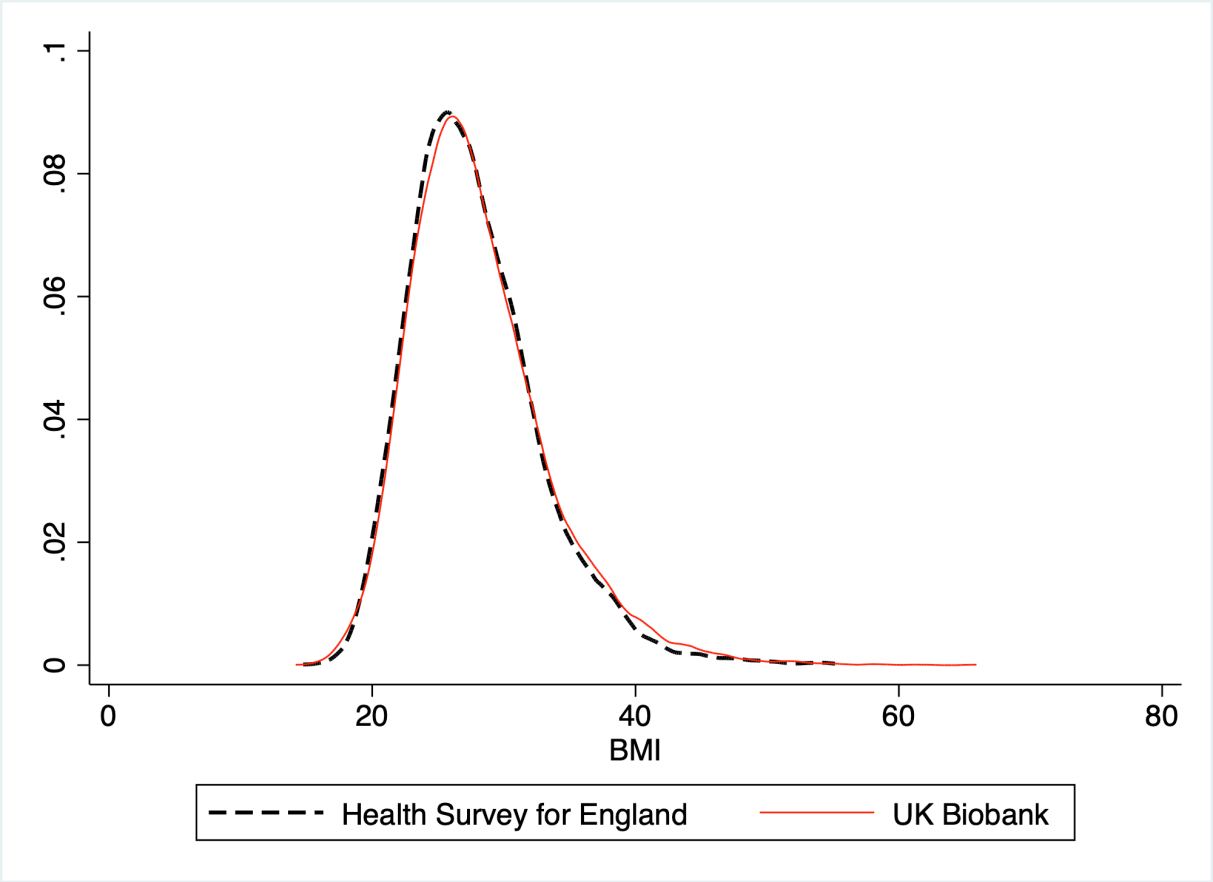
Appendix Table F4:
Effect on Smoking, Physical Activity, and Hypertension Diagnosis and Medication

	Currently Smoke	*Ever Smoke*	*Hypertension Diagnosis*	*Hypertension Medication*	*Physical Activity*
Stayed in school until 16	0.00	0.02	-0.04	0.01	0.01
	[0.027]	[0.042]	[0.033]	[0.022]	[2.699]
Mean of Y	0.118	0.396	0.205	0.073	28.88
N Observations	129,296	127,711	129,190	127,967	29,436

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

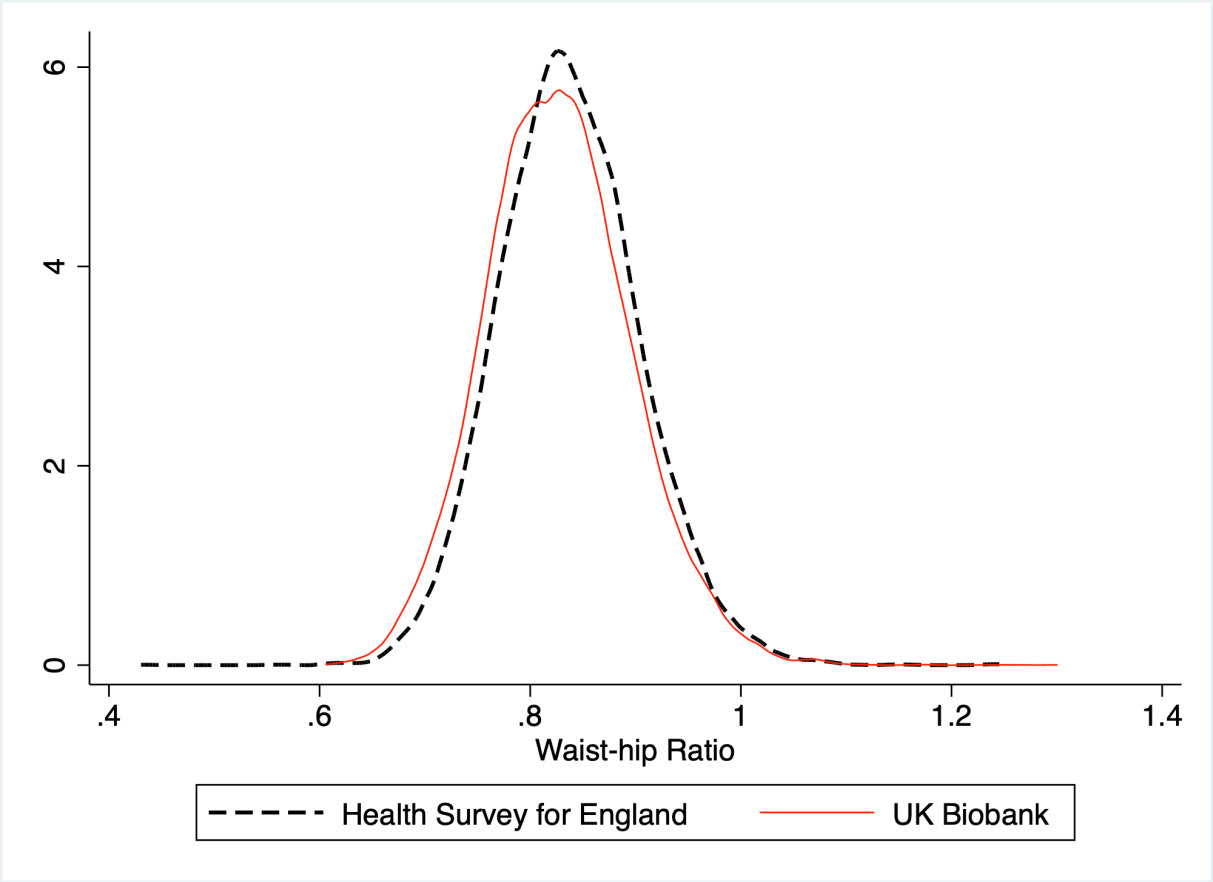
Appendix G

Appendix Figure G1: Comparison of Distribution of BMI of Compliers in the UK Biobank and in the Health Survey for England



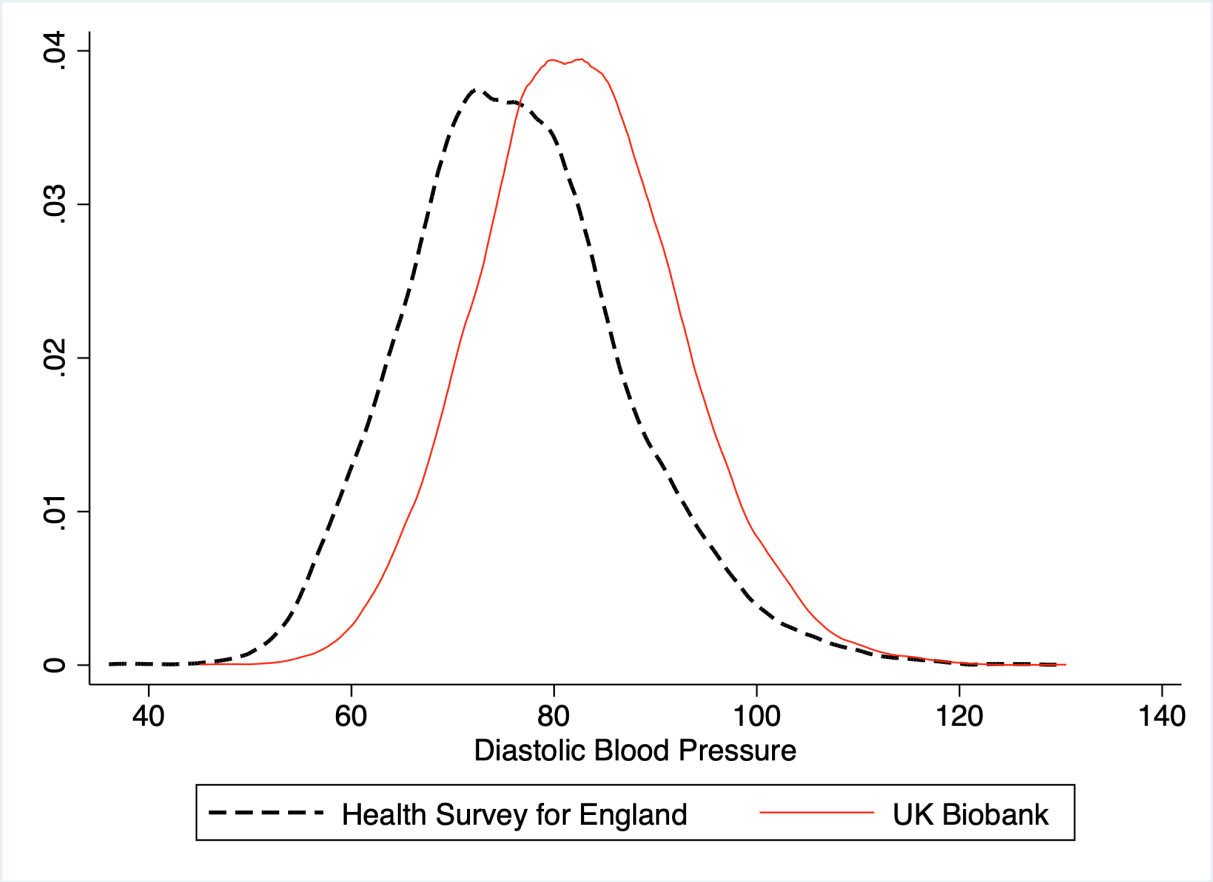
Notes: The figure compares the distribution of body mass index (BMI) of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents 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 = 13,438 (UKB), 8,450 (HSE).

Appendix Figure G2: Comparison of Distribution of Waist-Hip Ratio of Compliers in the UK Biobank and in the Health Survey for England



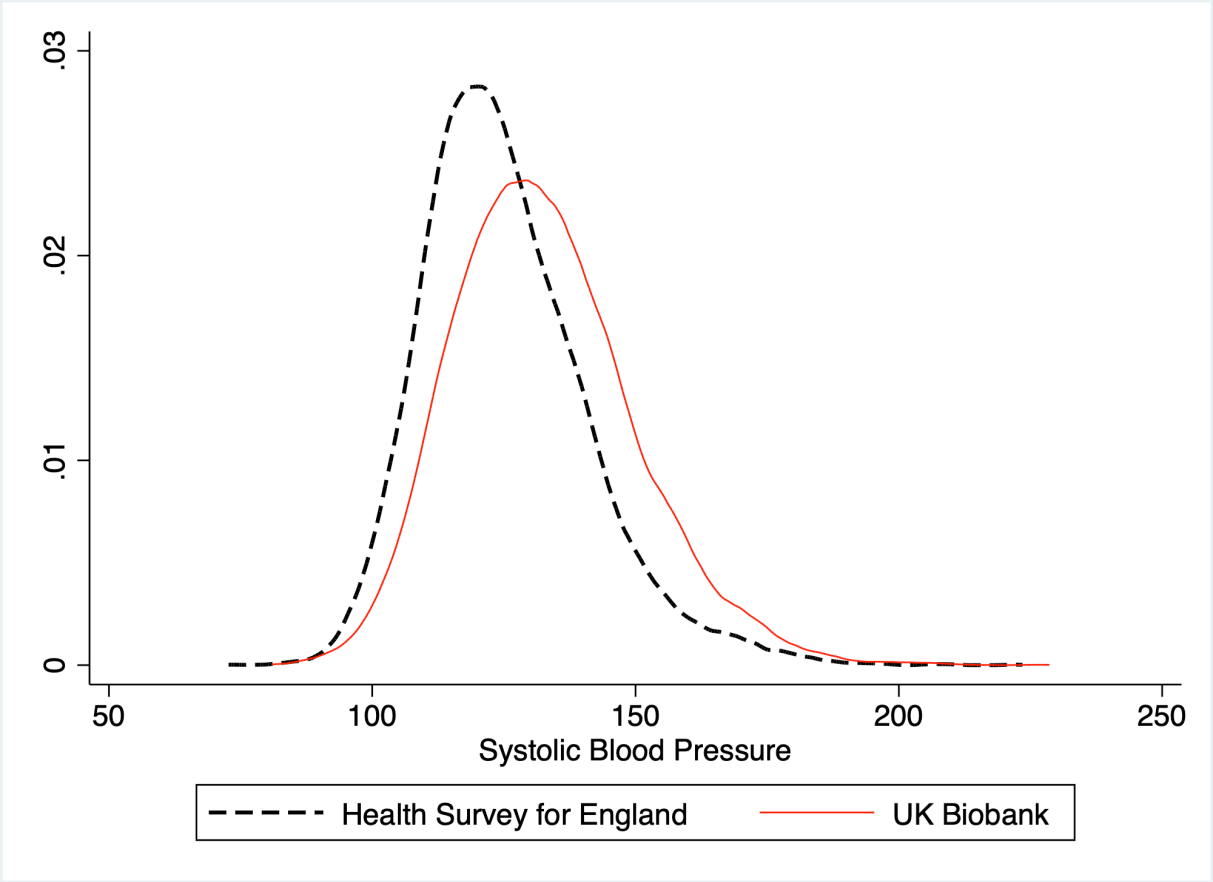
Notes: The figure compares the distribution of waist-hip ratio of compliers in the UK Biobank and in the Health Survey for England (years 2006, 2007, 2008, 2009, and 2010). The UK Biobank sample is restricted to respondents living in England. We do not have data on the date of birth of HSE respondents 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 = 13,460 (UKB), 6,972 (HSE).

Appendix Figure G3: 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 = 13,473 (UKB), 5,954 (HSE).

Appendix Figure G4: 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 = 13,473 (UKB), 5,954 (HSE).

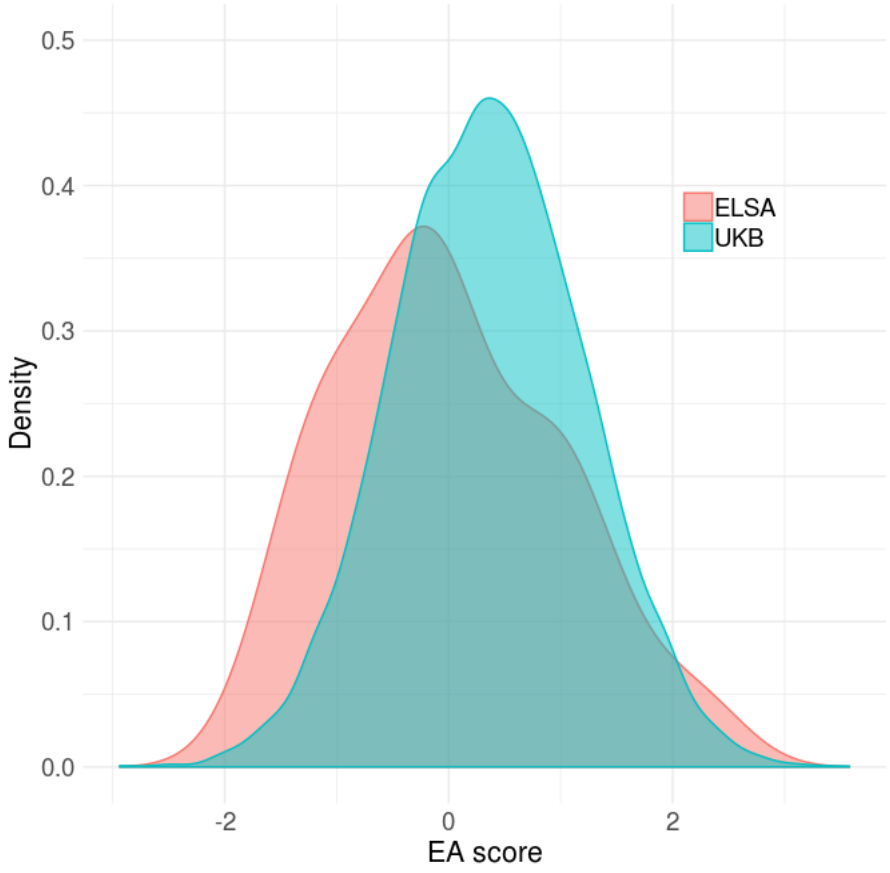
We used genetic data to re-weight the UK Biobank (UKB) sample in an attempt to make it nationally representative. Genetic data may be useful in this regard because it is fixed at conception. We used the English Longitudinal Survey of Ageing (ELSA) as a nationally representative benchmark. We restricted 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 G5: 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. Notice that the first stage estimates with and without weighting are largely similar and do not attain the level of the estimates observed in representative data (e.g., Clark and Royer, 2013). This implies that the PGS alone is not sufficient to explain the differences in representativeness between the UKB and other representative samples.

Appendix Table G1: Sensitivity of the Estimates of the Effects on Education to Re-weighting Based on Genetic Data

	Left School at Age 16 or Older			
Post	0.133	0.133	0.154	0.154
	[0.003]***	[0.003]***	[0.005]***	[0.005]***
Weighted	No	Yes	No	Yes
Controls?	No	No	Yes	Yes
Mean of Y	0.829	0.829	0.829	0.829

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 = 125,948. 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 H

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

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

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 I

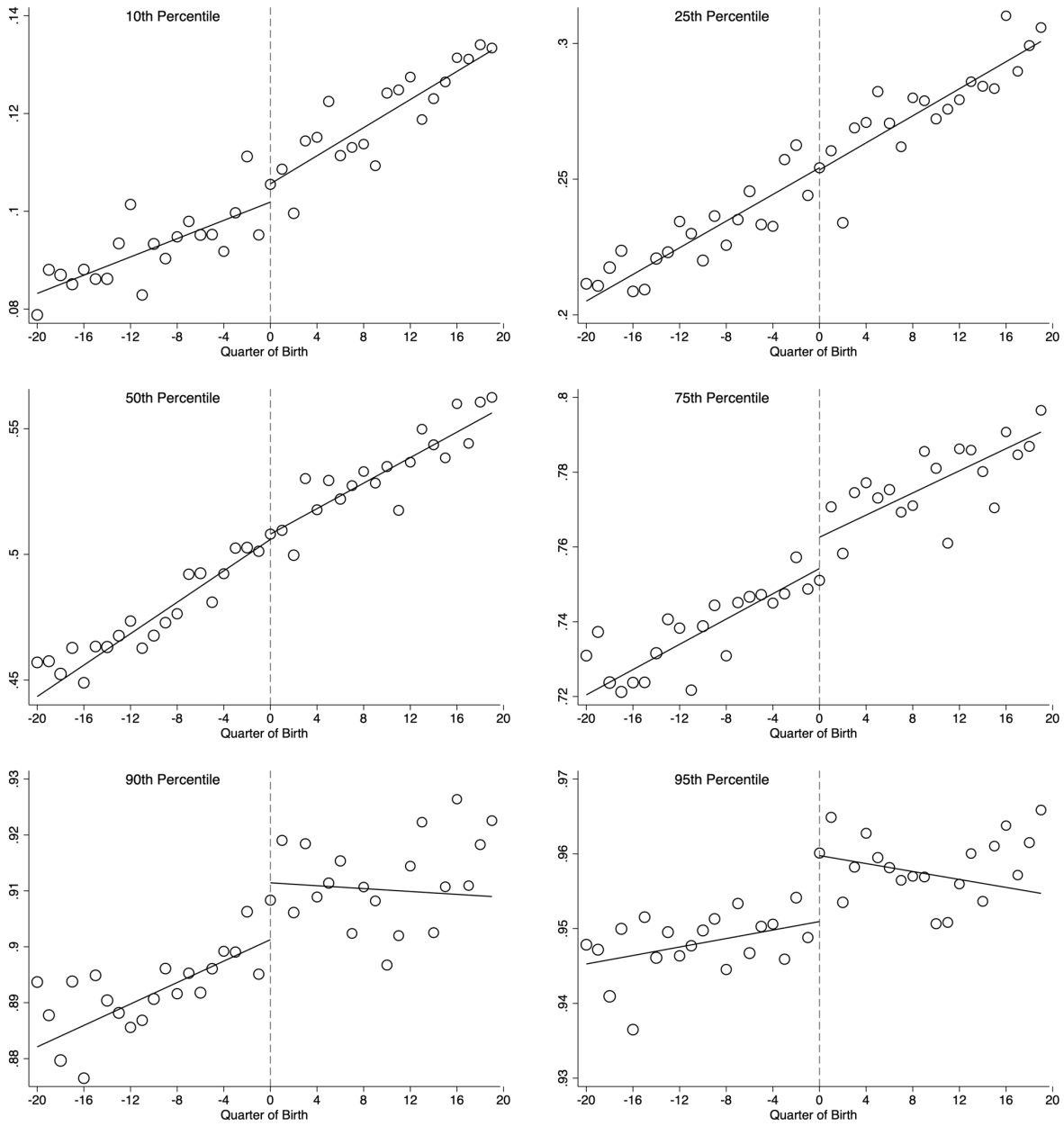
Appendix Table I1: Effects of 1972 ROSLA on Education

	Left School at Age 16 or Older		Left School at Age 17 or Older		Left School at Age 18 or Older		Left School at Age 19 or Older	
Post	0.136	0.136	0.005	0.005	0.007	0.007	-0.007	-0.007
	[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	129,370	129,370	129,370	129,370	129,370	129,370	129,370	129,370
Mean of Y	0.826	0.826	0.612	0.612	0.532	0.532	0.441	0.441

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.

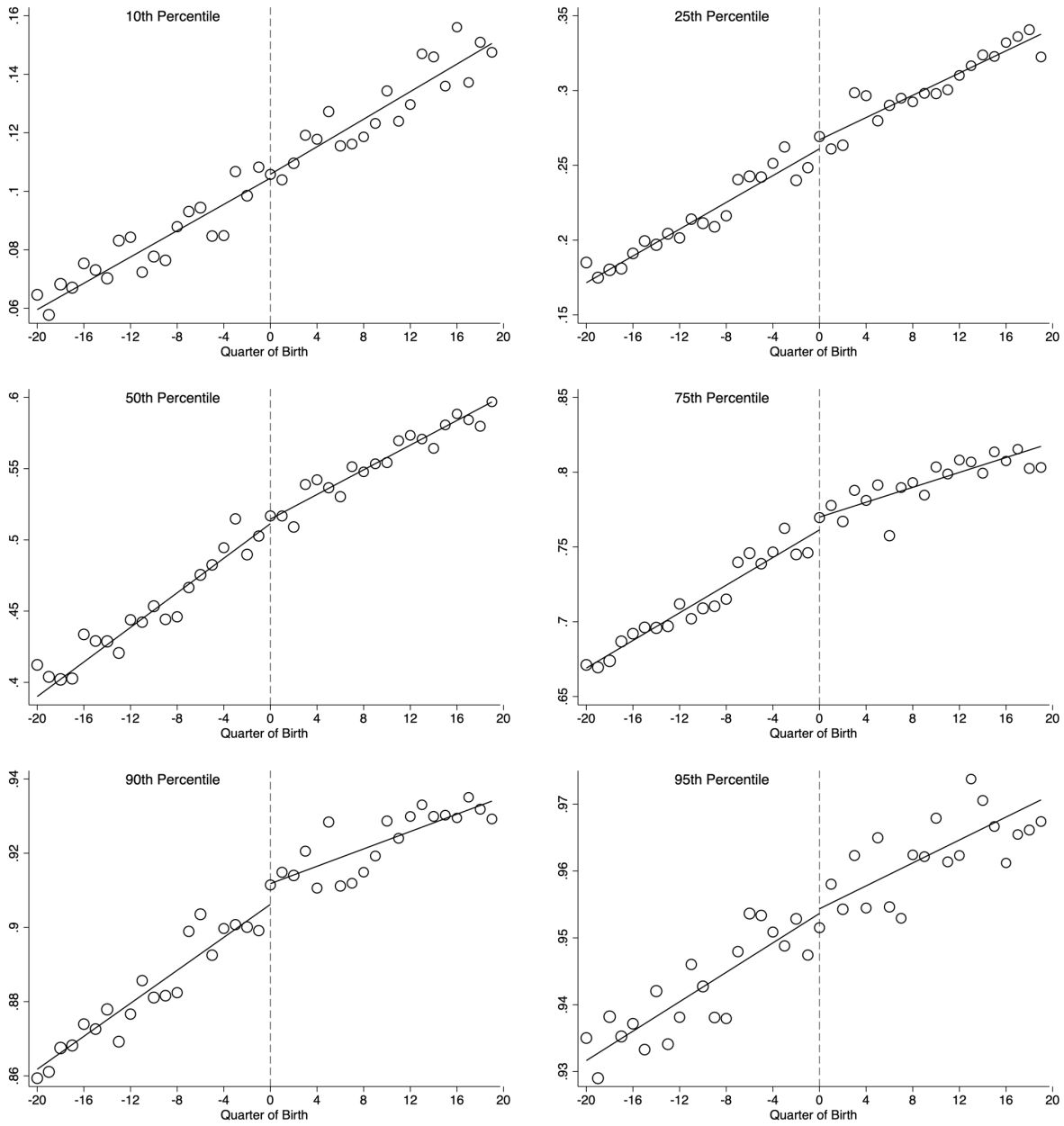
Appendix J

Appendix Figure J1: Effect on (Pre-ROSLA) Percentiles of Body Size Distribution



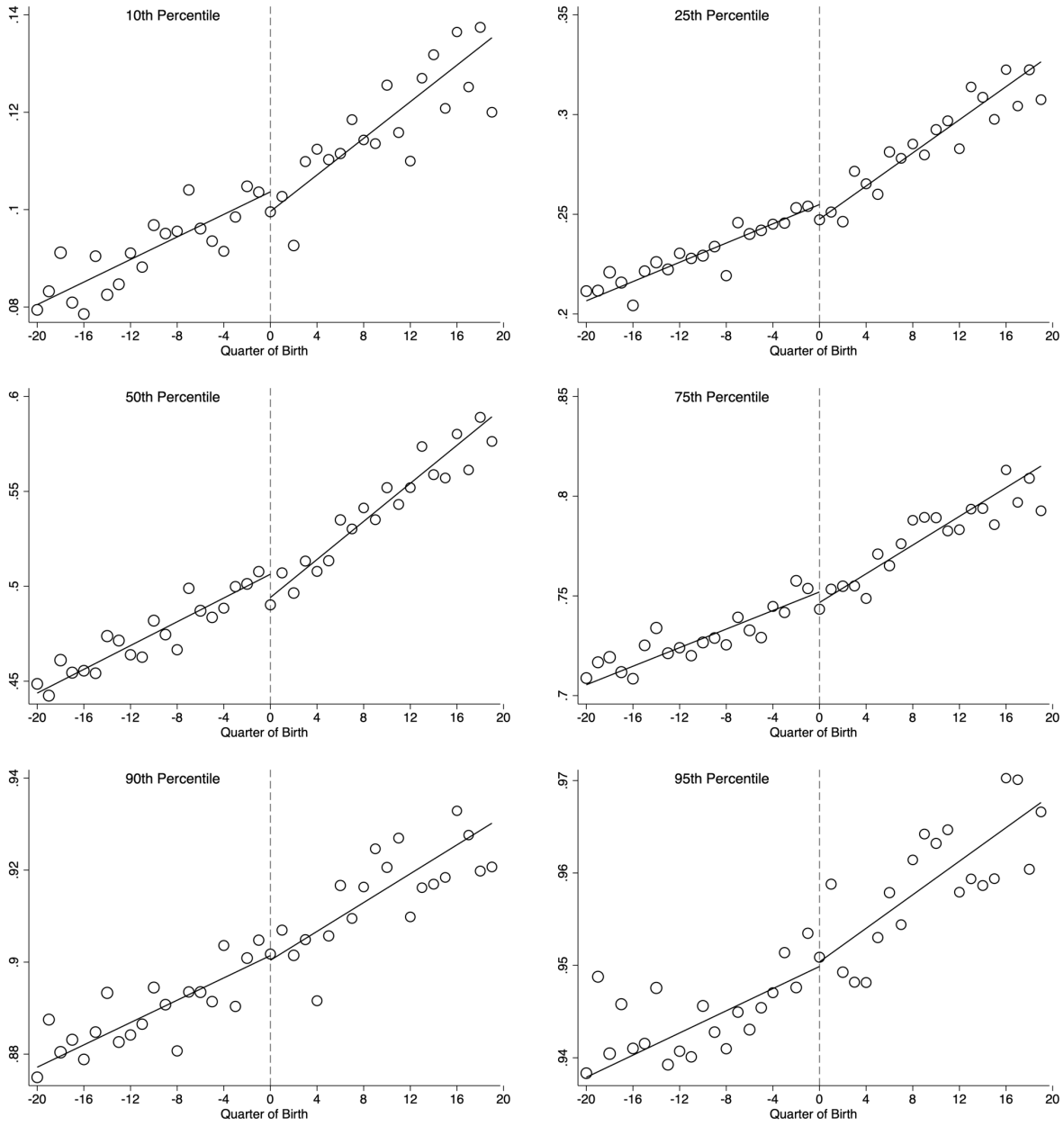
Notes: The figure shed lights on the results shown in Figure 4 of the paper. It shows the fraction of study participants with a body size index below the 10th, 25th, 50th, 75th, 90th, and the 95th percentile (of the distribution of those born between September 1, 1956 and August 31, 1957) by quarter 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 = 127,366.

Appendix Figure J2: Effect on (Pre-ROSLA) Percentiles of Lung Function Distribution



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 = 103,682.

Appendix Figure J3: Effect on (Pre-ROSLA) Percentiles of Blood Pressure Distribution



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 = 129,165

Appendix K

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	Number of obs	= 16
Residual	.031789143	13	.002445319	F(2, 13)	= 759.50
				Prob > F	= 0.0000
Total	3.74624377	15	.249749585	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, 28.46399, 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, –0.431959. The post-reform average BMI among compliers is equal to 28.032031= 28.46399 – 0.431959. Finally, the estimated reduction in mortality is $\left[\left(\frac{\widehat{RR}(28.032031)}{\widehat{RR}(28.46399)}\right) - 1\right] * 100$

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.