

No effect of additional education on long-term brain structure – a preregistered natural experiment in thousands of individuals

Nicholas Judd , Rogier Kievit

Cognitive Neuroscience Department, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

A regression discontinuity analysis finds essentially no effect of 1 additional year of secondary education on brain structure in adulthood. This is a **valuable** finding that adds to the literature on the impact of education on brain health. The evidence presented is **solid**, with strengths including methodological novelty as well as principled study design; the impact is, however, limited as the manipulated variable only relates to a single additional year of education (remaining in education to 15 vs 16 years of age). The interpretation is further missing discussion of the healthy volunteer bias of the UK Biobank sample, amplified in the imaging extension.

<https://doi.org/10.7554/eLife.101526.1.sa4>

Abstract

Education is related to a wide variety of beneficial health, behavioral, and societal outcomes. However, whether education causes long-term structural changes in the brain remains unclear. A pressing challenge is that individuals self-select into continued education, thereby introducing a wide variety of environmental and genetic confounders. Fortunately, natural experiments allow us to isolate the causal impact of increased education from individual (and societal) characteristics. Here, we exploit a policy change in the UK (the 1972 ROSLA act) that increased the amount of mandatory schooling from 15 to 16 years of age to study the impact of education on long-term structural brain outcomes in a large ($n \sim 30,000$, UK Biobank) sample. Using regression discontinuity – a causal inference method – we find no evidence of an effect from an additional year of education on any structural neuroimaging outcomes. This null result is robust across modalities, regions, and analysis strategies. An additional year of education is a substantial cognitive intervention, yet we find no evidence for sustained experience-dependent plasticity. Our results provide a challenge for prominent accounts of cognitive or ‘brain reserve’ theories which identify education as a major protective factor to lessen adverse aging effects. Our preregistered findings are one of the first implementations of regression discontinuity on neural data – opening the door for causal inference in population-based neuroimaging.

Introduction

Access to education is codified as a fundamental human right with immense societal and economic benefits ^{1,2}. Individuals who experience more education tend to show a wide variety of beneficial health, cognitive, and neural outcomes ³. Correlational evidence across disparate contexts, countries, and demographics offers overwhelming support for these findings ^{4–10}.

Education not only increases learned skills and knowledge, but also reasoning and fluid intelligence abilities ^{11–14}. One plausible mechanism underlying these widespread benefits from education is the long-term reorganization of brain structure. Lifespan theories on heterogenous neurodevelopment place particular emphasis on education ^{15–19} and correlative work further support the role of brain structure as a mechanism ^{7,20–22}. For instance, more educated individuals show higher mean cortical thickness in later life – taken as evidence of ‘increased brain reserve’ lessening the neurodegenerative effects of aging ^{23,24}.

Yet, strong causal evidence for the effect of education on brain structure does not yet exist. Both ethical and practical constraints mean that the effect of education cannot be experimentally tested. This makes it unclear if any findings from education are causal, or if instead, they reflect a complex web of preexisting, or amplifying, sociodemographic and individual characteristics ^{3,25} – many of which have been associated with brain structure (e.g., intelligence ²⁶, parental income ²⁷, neighborhood pollution ²⁸). Among the plethora of environmental factors, there is also a substantial (~40% heritability) genetic component implicated in educational attainment, further confounding potential effects ^{29,30}. In addition, access to higher education involves substantial selection processes at the level of the individual and educational system which further complicate the causal pathways involved.

One solution to this problem is using a *natural experiment* – a ‘random-like’ (exogenous) external event – allowing causal inference in observational phenomena ^{31,32}. A crucial feature of this design is an assignment mechanism outside a participant’s control, usually of natural (e.g., geographical, weather events) or governmental (e.g., policy decisions, cutoff rules) origin. For instance, a law that increases the number of mandatory school years affects everyone equally, in turn, decoupling individual characteristics and other unmeasured confounders from the effects of additional education ³³. One of the most widely used causal inference analysis techniques is regression discontinuity (RD), which is applicable when treatment is assigned via a cutoff of a particular score or running variable ^{31,32}. Recent advances in the field of econometrics have optimized inference and largely standardized regression discontinuity (RD) analysis ^{32,34–36}. These methods have provided robust evidence that additional education *causes* an increase in intelligence (0.14 – 0.35 SD) ¹². Despite the many strengths and potential causal insights natural experiments offer the field of cognitive neuroscience they have not been used to address such questions.

On September 1st 1972, the minimum mandatory age to leave school was raised from 15 to 16 years of age in England, Scotland, and Wales (called here ‘ROSLA’) ³⁷. The consequence of this law change was substantial: it resulted in almost 100% of children aged 15 staying in school for an additional year, in turn, increasing formal qualifications, income, and cognition ^{38–43}.

However, whether this substantial intervention also affected the long-term brain structure of those born right around the cut-off remains an open question. This is unfortunate, as regression discontinuity is a powerful tool to study phenomena that cannot (ethically or practically) be randomized. Recent developments of increased large population-based neuroimaging cohorts provide the sample size needed to make use of these cross-disciplinary methods. One such cohort,

the UK BioBank, has recruitment criteria matching the geographic and birth window of the natural experiment ROSLA ⁴⁴. This provides the ideal opportunity to test, for the first time, the causal effect of a year of education on long-term structural neuroimaging properties.

Using a preregistered design (<https://osf.io/rv38z/>) with over 30,000 participants we evaluate if an additional year of education, as mandated by ROSLA, causes changes in six global neural properties (total surface area, average cortical thickness, normalized total brain volume, mean weighted fractional anisotropy, white matter hyperintensities, and normalized cerebral spinal fluid volume) as compared to individuals born before the cut-off. It is also possible an effect from education could manifest only in specific regions, in the absence of a broad, macro-neural effects. For this reason, we further tested 66 individual cortical regions for surface area and cortical thickness, 27 white matter fiber tracks for fractional anisotropy, and subcortical gray matter volume in 18 regions. Taken together, the combination of cutting-edge quantitative methods, a large sample, a well-validated natural experiment and high-quality imaging allows us to examine, for the first time, if an additional year of education causes long-term structural reorganization in the brain.

Results

To test if an additional year of education causes long-lasting changes in neural properties we used fuzzy local-linear regression discontinuity (RD) ^{34,35}. This continuity-based technique exploits the fact that ROSLA affected individuals based on a date of birth cutoff (September 1st, 1957). More specifically, adolescents born after this date had to spend one more year in school than those born only one day earlier. Local-linear RD analysis tests this by comparing the limits of two non-parametric functions, one fit using only participants right before the policy change to another function fit on participants right after ³². If an additional year of education affects neural outcomes, we assume that brain structure will be *discontinuous* exactly at the cutoff. In contrast, if ROSLA does not affect long-term neural outcomes the functions should be *continuous*, or smooth around the cutoff.

Using fuzzy local-linear RD we fit a series of regressions to various measures of brain structure (Fig 1) to test for any discontinuity at the cutoff. To determine the optimal number of participants to include, we use mean square error optimized bandwidths to a maximal range of 10 years before and after the cutoff ($N > 30,000$). Doing so, we observed no evidence of an effect from additional education on any of our preregistered global neuroimaging measures (p 's $> .05$; Sup. Table 2). In other words, the relationship between the year of birth and neural outcomes was *continuous* around ROSLA's cutoff, indicating no differences in global structural measures from an extra year of education. The optimized bandwidths used in this analysis included participants born 20 to 35 months around the cutoff (average $N = 5124$). The absence of a causal effect of education was observed for all our global neuroimaging metrics: total surface area (SA), average cortical thickness (CT), normalized total brain volume (TBV), mean weighted fractional anisotropy (wFA), white matter hyperintensities (WMh), or normalized cerebral spinal fluid volume (CSF) (Figure 1; Sup. Fig. 3). These results did not change when imputing missing (~4%) covariate data.

These findings strongly suggest that an additional year of education did not lead to changes detectable by MRI decades later. However, to ensure the validity of our (causal) inferences, a critical step in any regression discontinuity approach is to test the validity of the design ³². For instance, if participants can manipulate their treatment by sorting around the cutoff this severely limits the strength of causal claims. In the context of ROSLA, this is highly unlikely since the assignment was based on date of birth ^{37,45}. However, for completeness we tested this question ⁴⁶, finding no evidence that individuals were somehow able to adjust their enrollment ($T_q = -0.72$, $p = .47$; Sup. Fig. 2). A second validation approach is to employ *placebo outcome tests*

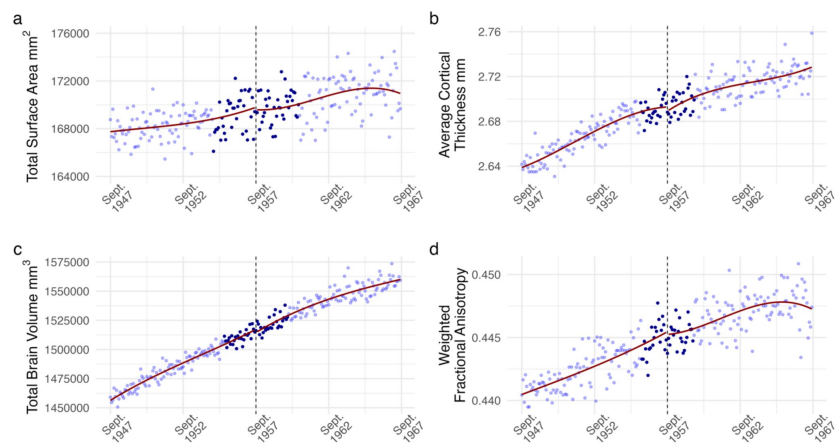


Figure 1

Regression discontinuity (RD) plot of monthly averaged *a)* total surface area, *b)* average cortical thickness, *c)* total brain volume, and *d)* weighted fractional anisotropy plotted by the participant's date of birth in months (our running variable). Each dot reflects the average value for individuals born in that month. The dashed line corresponds to Sept. 1957 the date of birth inclusion cutoff for an additional year of mandatory education from ROSLA. We found no evidence of an effect from an additional year of education on any structural neuroimaging measures – illustrated here by a continuous line around the cutoff. Dark blue dots represent all individuals within the mean-squared error-optimized bandwidths, in turn, reflecting participants used for analysis with a local-linear fuzzy RD approach. Third-order polynomials (dark red lines) are fit on either side of the cutoff *only* for illustration purposes. **Sup. Fig. 3** [illustrates the two other preregistered global neuroimaging outcomes](#) (total white matter hyperintensities and cerebral spinal fluid volume).

³²[\[link\]](#). This approach uses variables that should *not* be causally related to your treatment (e.g., an additional year of education), to ensure the absence of spurious effects arising through unknown mechanisms. To accomplish this, we used all of our neuroimaging covariates (e.g., sex, head motion, site) as placebo outcomes, under the assumption that ROSLA should not affect these variables. Other than one covariate (summer) which was deterministically related to ROSLA (therefore excluded), none of the placebo outcomes were related to ROSLA (**Sup. Table 1** [\[link\]](#)). Our findings add to the existing body of prior work ³⁸[\[link\]](#), ⁴⁰[\[link\]](#), ⁴¹[\[link\]](#), ⁴³[\[link\]](#) further solidifying ROSLA as a valid natural experiment.

Next, we tested whether there may be any regionally specific neuroimaging effects. It is possible that an additional year of education caused localized neural changes that are not picked up globally. As preregistered, we used the Desikan-Killiany cortical atlas ⁴⁷[\[link\]](#) to test the effect of an additional year of education on 33 bilateral regions for both cortical thickness (CT) and surface area (SA). These analyses included on average 5080 effective participants (N range = 3884 – 7771) for CT and 5392 participants (N range = 3739 – 8727) for SA. Despite these relatively high participant numbers, we did not find an additional year of education to cause changes in any regions for CT or SA (p 's_{FDR} > .05). This was also the case for weighted mean fractional anisotropy in all 27 tracks tested ⁴⁸[\[link\]](#) (p 's_{FDR} > .05; mean n = 4766). Lastly, we tested the volume of 18 subcortical regions, finding none to be related to an additional year of education (p 's_{FDR} > .05, mean n = 5174). To summarize, there was no evidence of an additional year of education affecting any regional neuroimaging measures with fuzzy local-linear regression discontinuity (**Sup. Figure 4** [\[link\]](#)).

Bayesian Local Randomization Robustness Analysis

As an additional preregistered robustness test, we used a slightly different approach, often referred to as 'local randomization', to analyze natural experiments. This approach works under the assumption that individuals close to the cutoff are exchangeable and similar except for the treatment (in our case an additional year of school) ³²[\[link\]](#), ⁴⁹[\[link\]](#). To implement this analysis, we compared participants born *exactly* right before the cutoff (August 1957; n ≈ 130) to those born right after the cutoff (September 1957; n ≈ 100). An additional benefit is that we implemented this analysis in a Bayesian framework, allowing us to more readily interpret the strength of evidence either in favor of the null or alternative hypothesis. Our preregistered default point null Bayes factors were too wide, arguably providing relatively strong evidence in favor of the null. Taking a more conservative approach, we report these analyses with a narrower normal prior (mean=0, sd=1). Doing so, we replicated and extended our findings from fuzzy local-linear RD analysis, observing *strong* evidence in favor of the null hypothesis for total surface area (BF₀₁ = 18.21), average cortical thickness (BF₀₁ = 15.09), total brain volume (BF₀₁ = 13.61), weighted fractional anisotropy (BF₀₁ = 11.63), white matter hyperintensities (BF₀₁ = 13.26), and cerebral spinal fluid volume (BF₀₁ = 14.50). In addition, we tested across a range of priors which did not meaningfully affect our inferences, as each showed evidence in favor of the null (**Sup. Table 4** [\[link\]](#) & **6** [\[link\]](#)).

The two quantitative RD approaches described here (local-linear and local randomization) have strengths and challenges similar to the widespread bias versus variance tradeoff. As the number of months on either side of the cutoff increases, bias is introduced as participants become less similar, yet at the same time, the sample size increases, thereby lowering the variance of the estimate. Our local randomization approach runs a negligible risk of bias, but at the cost of a relatively modest sample size: One could argue the 230 participants included in our 1-month window are too few for neuroimaging outcomes. To examine the consequences of this tradeoff, we therefore expanded the boundary to a 5-month window around September 1st, 1957 (n per group ≈ 600). This larger participant pool provided further evidence in support of the null hypothesis of education not affecting global neural measures (**Figure 2** [\[link\]](#) & **3**, **Sup. Table 6** [\[link\]](#)). Lastly, we repeated our placebo outcome tests for both a one- and five-month window local randomization analysis, finding no associations (**Sup Table 5** [\[link\]](#) & **7** [\[link\]](#)), demonstrating the robustness of the natural experiment ROSLA and our analysis approach.

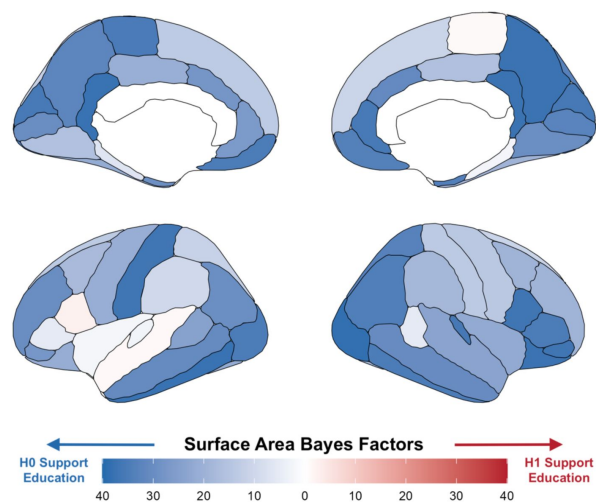


Figure 2

Bayes factors for surface area per region using a local randomization analysis with a 5-month window around the onset of ROSLA (September 1st, 1957). Illustrating widespread evidence against the effect of a year of education on total surface area. The regionally specific analysis of these bayes factors [reported prior: $\text{normal}(0, 1)$] was not preregistered and serves to illustrate our global neural findings.

Correlational Effect of Education

The above analyses employed an RD design allowing us to investigate the hypothesized causal effect of (additional) education on differences in brain structure independent of confounding pathways. To ensure our sample is at least *in principle* sensitive to observing brain-behavior associations (cf. [50](#)), we reran the analysis as a simple association instead. This allows us to examine whether more years of education are associated with differences in brain structure. Notably, such an observational analysis (and resulting parameter estimate) would reflect an indeterminate mixture of causal effects as well as any indirect, sociodemographic, and individual pathways. These associations would still be of considerable potential scientific interest, but could not be interpreted as *causal* effects of education on brain structure. Crucially, this was done using the same subset of participants as in the local randomization analysis. Resulting in an estimate of how much one additional year of education correlates with brain structure.

First, we estimated the association between education (in years of attainment) and global neuroimaging measures using the same sample of participants from the one-month window local randomization analysis (i.e., August & September 1957; $n \sim 230$). Similar to the causal approach, five of the six measures showed evidence in support of the null hypothesis (**Sup. Table 4**). In contrast, total surface area showed weak evidence in support of a positive association of education ($BF_{10} = 2.3$, $n = 229$). To increase power, we expanded our observational analysis to participants born in a five-month window around the ROSLA cutoff (**Fig. 3**). This considerably increased the strength of evidence in support of a positive association between years of education and total surface area ($BF_{10} = 41.7$, $N = 1185$). In addition, the larger pool of participants provided extreme evidence in support of an association between years of education and cerebral spinal fluid volume ($BF_{10} = 80.7$, $n = 1193$). This analysis highlights the stark difference between a causal and associational approach, while also providing evidence of sensitivity to brain-behavioral associations in the same sample. Of the remaining four global measures, three meaningful decreased in their strength of evidence (in terms of Jeffreys criteria⁵¹) in favor of the null when compared to the causal estimate (**Fig. 3**; **Sup. Table 6**). The only global neuroimaging measure that provided a similar amount of evidence (in favor of the null) between a causal and correlative approach was mean cortical thickness ($BF_{01} = 7.22$ & $BF_{01} = 8.81$ respectively).

Discussion

In a large, preregistered study we find converging evidence *against* a causal effect of education on long-term structural neuroimaging outcomes. This null result is present across imaging modalities, different regions, and analysis strategies. We find no issues with the design of the 1972 ROSLA, substantiating it as a valid natural experiment, in agreement with prior work [38](#), [40](#)–[43](#). Despite a large sample (min $N = 4238$), we find no evidence of an effect of education on any of the global neuroimaging measures with a continuity-based RD analysis. Confirming this result, we find strong evidence in support of the null hypothesis for these global neuroimaging measures using a Bayesian local randomization analysis. Moreover, we find no regionally specific effect of education on local mean cortical thicknesses or surface area across 66 cortical regions. This lack of localized effects was further confirmed in weighted mean fractional anisotropy for 27 white matter tracks, as well as subcortical gray matter volume in 18 regions. Moreover, we demonstrate the ability to find strong evidence in favor of observational associations between education and brain structure at this resolution, suggesting our findings are not due to lack of sensitivity more generally.

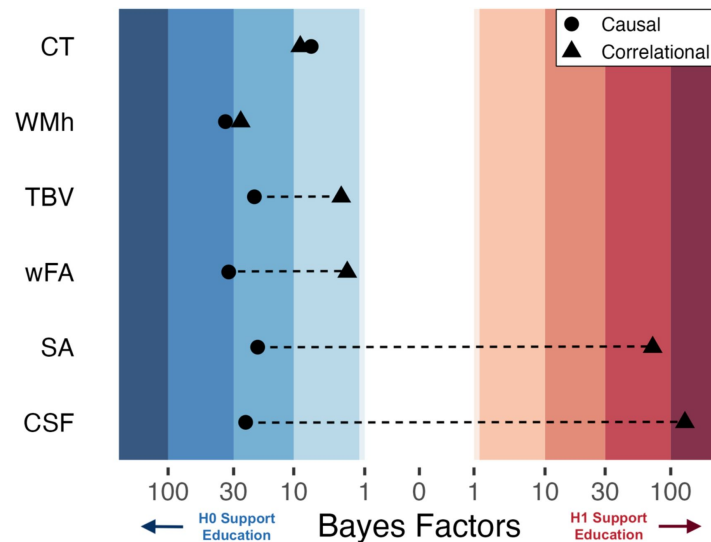


Figure 3

Bayesian Heat Plot:

Bayesian evidentiary strength (x-axis) for causal (circles) and correlational (triangles) estimates reflecting the impact of one additional year of education on global neuroimaging measures (y-axis; average cortical thickness [CT], white matter hyperintensities [WMh], total brain volume normalized for head size [TBV], mean weighted fractional anisotropy [wFA], total surface area [SA], and cerebral spinal fluid volume normalized for head size [CSF]). Stripped bands reflect the strength of evidence using Jeffrey's criteria 1961. For the causal estimates, positive Bayes Factors indicated support for the alternative hypothesis that an additional year of education affects the brain, while negative values indicate support for the null hypothesis of no effect. The causal and correlational parameters come from the same set of participants ($n \approx 1200$) born from April 1957 until Jan 1958. The causal parameter is an estimate of the effect of ROSLA with a 5-month window local-randomization analysis, the correlational parameter is an estimate of the association between a participant's self-reported educational attainment in years and (global) neuroimaging measures.

Our robust null result is seemingly at odds with causal inference findings of education's positive behavioral effect on intelligence ¹¹ [14](#) [39](#) – which is sustained throughout decades ¹² [38](#). This juxtaposition suggests that, to the extent that the additional year of education induced long-term changes in cognitive abilities, the neural manifestation is at a level of resolution not detectable with conventional MRI field strengths. However, this would seem to contrast with a range of influential findings demonstrating that high-intensity experimental behavioral interventions (e.g. juggling, studying, memory training) lead to measurable differences in brain structure (with similar imaging pipelines) in much smaller samples ⁵² [54](#). Moreover, compared to even these high-intensity interventions, a year of education is an *extensive* period of learning. The 1972 ROSLA was well implemented, schools had time to prepare and were given additional funding, increasing standardized formal qualifications of those affected ⁴¹ [42](#). This leaves open the question of how to interpret this constellation of findings.

One potential explanation to account for this discrepancy is the concept of *expansion-renormalization* ⁵⁵ [56](#), which posits following a period of skill acquisition, the cortex initially expands and then renormalizes over the course of a few months. In our context, this would suggest that the additional year would have manifested at a level detectable in MRI when the difference in educational exposure between children pre- and post-ROSLA was most pronounced and recent. In other words, MRI effects at the macro-scale might have been detectable immediately post-ROSLA in 16-year-old adolescents, before renormalizing, to a micro-scale, leaving in place permanent, but microstructural changes. Possible cellular candidates for initial experience-dependent plasticity are an increase in dendritic spines, the swelling of astrocytes, and intracortical myelin adaptations ⁵⁷ [58](#). These structural changes may be detectable using other approaches such as in vivo cellular work (cf. ⁵⁹ [60](#)), extreme high field strengths ⁵⁶ [60](#), or postmortem histology ⁶¹ [62](#).

Additionally, the long period of time between additional education and neuroimaging offers both strengths and weaknesses for our design. First, it could be the case that 46 years is too long and any potential effect faded out over the years. That is, rather than having a micro-neural effect, it may be that there simply are no lingering effects at the brain level at all. In this case, it may be better to think of the (causal) impact of additional education as more akin to fitness or strength interventions which are also unlikely to persist across such a period. However, we note that prominent aging-related theories of heterogeneity argue *directly against* this rationale, instead positing life course experiences offer a reserve or 'brain buffer' that leads to an increasing cascade of processes limiting adverse aging effects ¹⁵ [17](#). Here, the timing between our intervention (ROSLA) and scanning makes our design particularly well-powered to test these theories, since education would contribute to an initial brain buffer (intercept) and any cumulative educational effect (slope) over 46 years. While our results are at odds with prior conceptual and observational work, a recent longitudinal study found prior education to not affect the rate of brain aging ⁶³ [64](#) – in alignment with our findings.

Lastly, to demonstrate the importance of controlling for unobserved confounders we conducted a simple correlation analysis – using the same subset of participants – relating education to differences in brain structure. We found evidence to support an *association* between more years of education and greater surface area and cerebral spinal fluid volume. This result emphasizes the need for caution in attributing causation in non-causal designs, as unobserved confounders can masquerade as an effect of interest. For instance, more education is frequently highlighted as offering behavioral and neural protection against the adverse effects of aging ¹⁵ [17](#) – while we replicate this inference associationally, we find no causal evidence for any neuroprotective effects. This suggests a more complex pathway of effects unfolding over time. Environmental causes are most likely very small and additive, which makes them not only difficult to study ⁶⁴ [65](#) but also equally hard to adequately control ⁶⁵ [66](#). Our findings suggest that to truly understand the neural and behavioral processes that unfold after interventions such as education we need a multipronged, mixed methods approach that combines deep phenotyping, longitudinal imaging

and behavioral follow-up ⁶⁶[66](#)–⁶⁸[68](#), as well as more sophisticated models that can capture gene-by-environment-interplay ¹⁴[14](#),⁶⁴[64](#),⁶⁹[69](#). Only then will we be able to identify idiosyncratic environmental effects and individual characteristics underlying heterogeneous lifespan development.

Here, we report a lack of causal evidence of a year of school on long-term neural outcomes in thousands of participants. An additional year of education is a substantial intervention and our preregistered findings are robust across imaging modalities, different regions, and analysis strategies. While our design cannot inform us of any short-term neural effects of education, our results call into question *sustained* experience-dependent plasticity, with significant ramifications for prominent theories of aging-related heterogeneity. The recent availability of large neuroimaging cohort data paired with cutting-edge methods from econometrics offers new avenues in studying neural effects. Causal inference is a new tool to the neuroimager's toolkit – opening novel, societally relevant phenomena – with the potential to move the field of population neuroscience from one of *association* to one of *causation*.

Methods

On September 1st, 1972 the minimum age to leave school was increased from 15 years of age to 16 in England, Scotland, and Wales ³⁷[37](#). This law, henceforth ROSLA, mandated children born after September 1st, 1957 to stay in school for an additional year. In contrast, a child born only one day earlier was unaffected and legally allowed to stop formal schooling at 15 years of age. The consequence of this law change was substantial: it resulted in almost 100% of children aged 15 staying in school for an additional year, in turn, increasing formal qualifications (**Sup. Fig 2** ⁴⁰[40](#),⁴¹[41](#)). Crucially for our purposes, ROSLA is a well-studied natural experiment with (commonly agreed, e.g. ³⁸[38](#),⁴⁰[40](#)–⁴²[42](#)) high design validity. A sizable body of prior work has found behavioral effects from the 1972 ROSLA ³⁸[38](#),⁴⁰[40](#)–⁴³[43](#).

In this study, we leverage ROSLA to study the causal effect of an additional year of education on the brain. To do so, we will use the neuroimaging sub-sample of the UK BioBank – the largest neuroimaging study to date – which also lines up perfectly with geographic and birth window (~1935-1971) requirement characteristics for ROSLA ⁷⁰[70](#). We followed our preregistration (<https://osf.io/rv38z> ⁴[4\) closely, yet some minor deviations were necessary and explicitly outlined in ‘deviations from preregistration’ \(code: <https://github.com/njudd/eduBrain> ⁴\[4\]\(#\)\).](#)

Structural neuroimaging outcomes

It is very plausible that a broad intervention like education could affect the brain in either a global, or more regionally specific manner. For this reason, we examined both whole-brain averaged measures in addition to regional specificity in atlas-based regions and tracts. We tested the following global measures; total surface area (SA), average cortical thickness (CT), total brain volume normalized for head size (TBV), mean weighted fractional anisotropy (wFA), white matter hyperintensities (WMh), and cerebral spinal fluid volume normalized for head size (CSF). Next, we examined cortical thickness (CT) and surface area (SA) regionally using the Desikan-Killiany Atlas ⁴⁷[47](#). The temporal pole was not included by the UK BioBank making the total number of regions 66. We will also test weighted mean fractional anisotropy (wFA) with a global average and, regionally on 27 white matter tracks ⁴⁸[48](#). Lastly, we examine subcortical volume in 18 regions ⁷¹[71](#).

All measures are derived from the image preprocessing pipeline from the UKB, further preprocessing details are outlined elsewhere ⁴⁴[44](#),⁷²[72](#). All outliers were identified if they were either above quartile 3 plus 1.5 times the interquartile range (IQR) or below quartile 1 minus 1.5 times the IQR and brought to the fence by manually recoding them to this limit (Tukey/Boxplot

Method). Our alpha level for global neuroimaging measures (mean CT, total SA, average FA, WM hyper-intensities, normalized TBV & normalized CSF) is 0.05. The regional metrics (SA, CT, wFA & subcortical structures) are false discovery rate (FDR) corrected using the number of regions per modality (e.g., 66 regions for SA) with a q value less than 0.05 considered significant. Neuroimaging measures are reported in raw units.

Continuity-based framework: Local-linear fuzzy regression discontinuity

Regression discontinuity (RD) is a technique we use to estimate the effect of an intervention on an outcome where assignment (usually binary) is based on a cutoff of a running or ‘forcing’ variable ^{31,32} – in our case, age in months. One major design issue in RD is if participants select into (or out of) the treatment group by sorting around to the cutoff of the running variable. However, as the 1972 ROSLA law was not pre-announced, generally strictly enforced, and affected teenagers, such alternative explanations are highly improbable ⁴⁵. Nevertheless, we conducted a density test of the running variable to check for bunching near the cutoff ⁴⁶. Although the design validity of the 1972 ROSLA is well established ^{38,40,43} we still tested a variety of placebo outcomes (outlined in ‘covariates of no interest’) – outcomes implausible to be affected by the intervention.

RD analysis broadly falls into two separate but complementary frameworks ^{36,49}. The first, continuity-based approach defines the estimand as the difference between the limits of two continuous non-parametric functions: one fit using only participants right before the policy change to another function fit on participants right after. The other, so-called *local randomization* approaches assume participants are ‘as if random’ in a small window (ω) around the cutoff (described more in detail in the section “Bayesian Local Randomization analysis”). In this case, the estimand is the mean group difference between participants before and after the cutoff within ω . As ω approaches zero around the cutoff, the estimand becomes conceptually more similar to continuity-based approaches.

To empirically test whether an additional year of education caused long-lasting global and regional neural changes we used a fuzzy local linear regression discontinuity design (a continuity-based approach) with robust confidence intervals from the RDHonest package ^{31,34,35}. Our outcome variables are the neuroimaging metrics described above, which were adjusted to increase statistical precision (see section ‘covariates of no interest’). The running variable (X) is a participant’s date of birth in months (mDOB), as is convention, it was centered at zero around the birth cutoff of ROSLA (September 1st, 1957). Our first-stage (fuzzy) outcome was a dummy coded variable of whether the participant completed at least 16 years of education. Participants who indicated they completed college were not asked this question, therefore we recorded their response as 21 years ³⁸. Our choice of using the RDHonest package was primarily due to its ability to provide accurate inference with discrete running variables (in our case mDOB). We used the default settings of local-linear analysis on MSE-derived bandwidths with triangular kernels ^{34,35}.

We included participants born in England, Scotland, or Wales 10 years on either side of the September 1st, 1957 cutoff (dob Sept. 1st, 1947 – Aug 31st, 1967) with neuroimaging data. The range of the running variable (age in our case) to include on either side of the cutoff (known as the bandwidth; z) is one of the most consequential analytical decisions in regression discontinuity designs. Prior work using ROSLA in the UK BioBank has analyzed bandwidths as large as 10 years ⁴³ to as small as a year ³⁸. Large bandwidths include more participants (decreasing variance) yet these participants are also further away from the cutoff, in turn, potentially increasing bias in the estimand ^{36,49}. Conversely, smaller bandwidths provide a less biased, yet noisier estimates. State-of-the-art continuity-based RD methods use data-driven bandwidth estimation such as mean squared error (MSE) optimized bandwidths ⁷³. Since we used this approach the

optimal bandwidth range and, in turn, the number of included participants, will differ per fitted model. This also makes it logically incompatible to sensitivity test our bandwidths ³⁶, since they are mean squared error derived – widening them will lower variance and, in turn, increase bias while tightening them will have the opposite effect. Lastly, the use of triangular kernels means participants further away from the cutoff will be weighted less than those closer ³⁶. Both MSE-optimized bandwidths and triangular kernels determine the number of ‘effective observations’ to be fit by a fuzzy local linear RD model.

For our global continuity-based analysis we made sure the results did not change due to missing covariate data. This was accomplished by imputing missing covariate data ($\approx 4\%$) with classification and regression trees from the MICE package ⁷⁴. We imputed using information from only the variables included in each analysis and did not use the running variable (DOB in months) or our first-stage instrument for prediction. We did this ten times checking the estimates and inferences across each iteration to ensure robustness.

Bayesian Local Randomization Analysis

As a robustness test and to provide evidence for the null hypothesis, we conducted a Bayesian analysis using the local randomization framework for RD ³². This alternative framework assumes a small window (ω) around the cutoff (c) where the running variable is treated “as if random” ⁴⁹. While the local randomization framework invokes stricter assumptions on the assignment mechanism, placing more importance on the window around the cutoff ($\omega = [c - w, c + w]$) it handles discrete running variables well ⁴⁹. As the number of months on either side of the cut-off increases, bias is introduced as subjects become less similar, yet at the same time, the sample size increases, thereby lowering the variance of the estimate.

As recommended ^{32,49}, we included participants within the smallest window possible ($\omega = 1$ month; August vs September 1957), then expanded to a 5-month window around September 1st, 1957. A dummy variable (ROSLA) was constructed to reflect if a participant was impacted by the policy change. We then tested the effect of this variable on our six global neuroimaging measures while correcting for the covariates listed above. A few neuroimaging covariates did not have sufficient observations to be included (see ‘deviations from preregistration’).

Models were fit in R (v. 4.3.2) with rstanarm with Markov Chain Monte Carlo sampling of 80,000 iterations over 4 chains ⁷⁵. All priors (p) used a normal distribution centered at 0 with autoscaling [$p \cdot \text{sd}(y) / \text{sd}(x)$]. Our preregistration referred to using the ‘default’ weakly informative prior of STAN (i.e., 2.5 SDs). However, this is a relatively wide prior for point null Bayesian hypothesis testing, and at odds with the defaults from packages meant for this purpose (e.g., BayesFactor). We therefore deviated from our preregistration and reported Bayes Factors with a normal prior centered at 0 with a standard deviation of 1 (medium informative). We also report strongly informative (SD = .5) and weakly informative (SD = 1.5) normal priors.

Model diagnostics were checked with trace plots, posterior distributions, and rhat values (< 1.05) ⁷⁶. We then computed log Bayes Factors using the bayestestR package ⁷⁷ using the Savage-Dickey density ratio with a point-null of 0 for each of the 3 priors. The strength of evidence was interpreted on a graded scale using the criteria preregistered ⁵¹. If the two frameworks disagree our primary inference will be based on the continuity-based framework.

Similarly to our continuity-based approach, we conducted placebo outcome tests within both windows ($\omega = 1$ & 5 months) using each included covariate as the outcome being predicted by ROSLA. In addition, we preregistered four placebo cutoffs – an analysis where the cutoff is artificially moved to test the specificity of the effect – yet null findings made this test no longer necessary (see ‘deviations from preregistration’).

Correlational Analysis

To compare our results to a correlational approach, we tested self-reported years of total education (EduYears) on the six global neuroimaging measures using the same participants as the local randomization analysis ($\omega = 1$ & 5 months). The analysis pipeline is identical to the local randomization approach, except the dummy coded term ROSLA was substituted for continuous EduYears.

Covariates of no interest

In neuroimaging it is common to include covariates. However, for identification in a valid regression discontinuity design covariates *by definition* should be equal on either side of the cutoff ³⁶. We used standard neuroimaging covariates for two purposes, 1) to further test the validity of the 1972 ROSLA and 2) to increase statistical precision in our RD analysis. For instance, there are large sex differences in certain neural measures mostly related to head size ⁷⁸, therefore we included a dummy variable for sex. We expect ROSLA to not affect the proportion of males and females, yet including this measure as a covariate will increase the estimation precision for measures sensitive to sex-related head size differences (e.g., surface area). [As expected, we found a 24% precision gain for total Surface Area, measured as the difference in CI width of the RD parameter pre and post-correction. Precision increased for *all* global measures (3-7%, 24%).]

Children born in July and August could technically leave school at 15 after 1972 ROSLA due to the exam period ending in mid-June and the school year starting on September 1st ⁴⁰. This is an artifact of asking ‘*What age did you complete full-time education?*’, rather than ‘*How many years of education did you complete?*’. This artifact predates and is not influenced by ROSLA ⁷⁹, yet to account for this we included a dummy variable for children born in July or August.

Lastly, Neuro-UKB recruitment is still ongoing and started in April 2014 ⁷⁰. In turn, there is a wide range of intervals in terms of the period between when someone was born and the age at which they were scanned. We control for this discrepancy by including a variable ‘date of scanning’ (DoS). This was done by coding the earliest date a (included) participant was scanned to 0 and counting the number of days until the last included participant. This resulted in a value for each participant that was the number of days they were scanned after the first participant. We also include this variable squared, so we can model quadratic effects, resulting in two variables (DoS & DoS²). Other (preregistered) neuroimaging quality control measures were included such as scanning site, head motion, whether a T2 FLAIR sequence was used, T1 intensity scaling, and 3 diffusion measures recommended ⁸⁰. No additional covariates other than those preregistered (<https://osf.io/rv38z>) were added.

Our first aim, known as a placebo outcome test, involved testing the effect of an additional year of education on each covariate. This was accomplished identically to our neuroimaging outcomes (see section ‘Continuity-based framework’) using RDHonest. We expect there to be no effect of ROSLA on any covariate. This was also tested for the Local-Randomization Framework ($\omega = 1$ & 5 months).

For our second aim, increasing the precision of the RD estimand we used covariate-adjusted outcomes (Y). This was done by fitting a local linear regression for each MSE-derived bandwidth z with a matrix of covariates (C) on the unadjusted outcome (Y) using Equation 1 below. Let X denote the running variable (date of birth in months) which is centered around the ROSLA cutoff (i.e., September 1st, 1957 = 0). In the second step, this matrix of covariates is then multiplied by the fitted coefficients (β) and subtracted from the unadjusted outcome (Y) to make a covariate-adjusted outcome (Y). Since our assignment mechanism is probabilistic (i.e., fuzzy RD) we also corrected our first-stage outcome, a dummy coded variable reflecting if the participant stayed in

school until 16, in an identical manner. This method was used after personal communication with Prof. Michal Kolesar the package creator of RDHonest. For the Bayesian local randomization analysis and correlational analysis, we simply included the covariates in the model.

$$uY = \beta_0 + \beta_1(X) + \beta_2(X \geq 0) + \beta_3(X * X \geq 0) + \beta_4 C' + \epsilon$$

$$Y = uY - (\beta_4 \times C')$$

Deviations from the preregistration

The study closely followed the preregistration pipeline (<https://osf.io/rv38z>), yet in a few minor cases it was not possible to follow. For instance, our initial plan involved placebo cutoffs – a sensitivity test where the cutoff is artificially moved to check if a significant result may also appear (for whatever reason) at other, non-hypothesized dates. Due to our lack of findings, we did not conduct this analysis as it was not necessary.

Some covariates lacked variance in specific specifications due to very few observations and therefore could not be included. In the global continuity analysis, this only impacted weighted fractional anisotropy, where the variable ‘T2_FLAIR’ was not used. This dummy-coded variable, indicating if a participant had a T2-weighted MRI scan, was also not used in the regional continuity-based analysis nor for the local-randomization analysis. The local randomization analysis included a small number of participants around September 1st 1957 therefore we did not include the covariate “summer” as this would have been isomorphic to our effect of interest (ROSLA). Lastly, for the one-month window analysis, there were not enough observations to include imaging center 11028.

Choice of priors

The Savage-Dickey density ratio is the height of the posterior divided by the height of the prior at a particular point (0 in our case for Point null Bayes Factors). This makes them particularly sensitive to the prior used. Our preregistration referred to using a ‘*default weakly informative prior*’. While not specified this was referencing the default prior of STAN (i.e., normal of 2.5 SDs). However, this is arguably too wide for adequate point null Bayesian hypothesis testing and at odds with the defaults from packages meant for this purpose (e.g., BayesFactor package). If we used the default prior from STAN it would have given us unrealistically strong support for the null hypothesis. We therefore deviated from our preregistration and reported Bayes Factors with a normal prior centered at 0 with a standard deviation of 1 (mediumly informative). We also report strongly informative (SD = .5) and weakly informative (SD = 1.5) normal priors both also centered at 0. All of our priors supported the null hypothesis they just varied in the amount of evidence in support of the null. Lastly, for illustration purposes, we ran the 5-month local randomization analysis for surface area regions using the mediumly informative prior (**Figure 2**).

Code and Data Availability

All code is publicly available (<https://github.com/njudd/eduBrain>). The data is also publicly available yet must be accessed via the centralized UK Biobank repository (<https://www.ukbiobank.ac.uk>). This research has been conducted using the UK Biobank resource under application number 23668.

Funding & Acknowledgments

Kievit was supported by a Hypatia fellowship from the RadboudUMC. This research has been conducted using the UK Biobank resource under application number 23668.

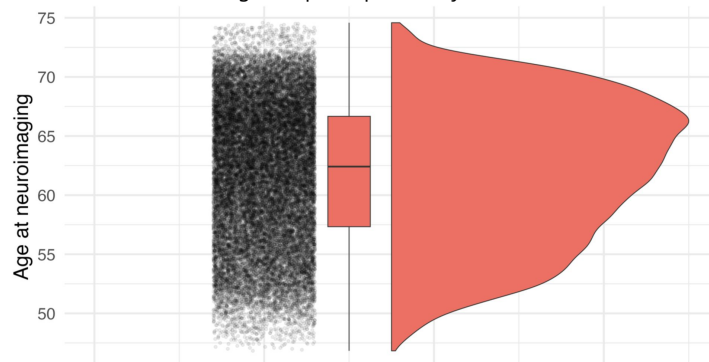
Ethical Statement

UK Biobank has ethical approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) (approval number: 11/NW/0382). This consortium has its own independent ethics advisory committee (<https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/governance/ethics-advisory-committee> ) , that assures the UK Biobank abiders by the ethics and governance framework (<https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf> ). Written informed consent was obtained from all participants.

Supplementary Information

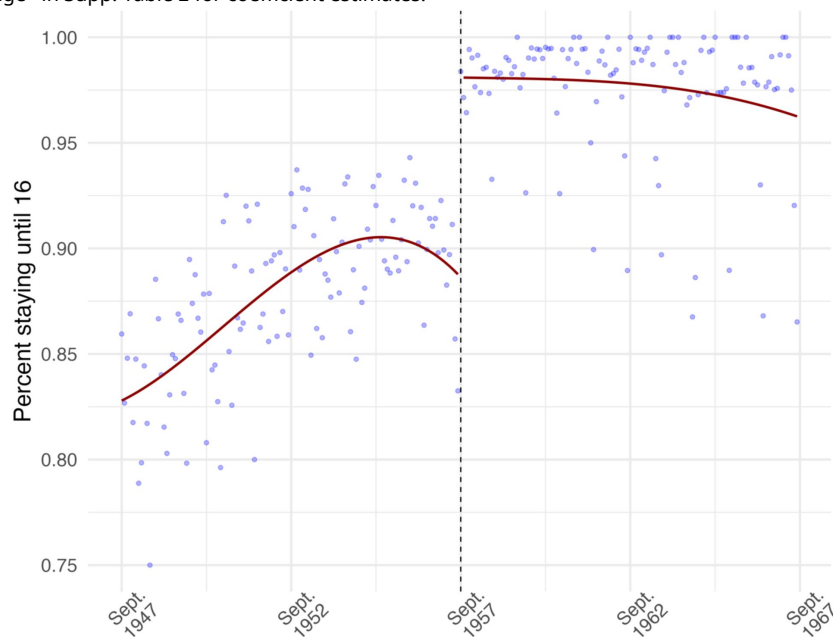
Supplementary Figure 1

Raincloud plot of the participant's age at neuroimaging (mean = 61.89). For anonymity reasons date of birth (DOB) is measured in months, yet the scan date was measured to the day. To derive age at neuroimaging we subtracted the number of months between DOB and the scan date setting each participant's day of birth to the first of the month.



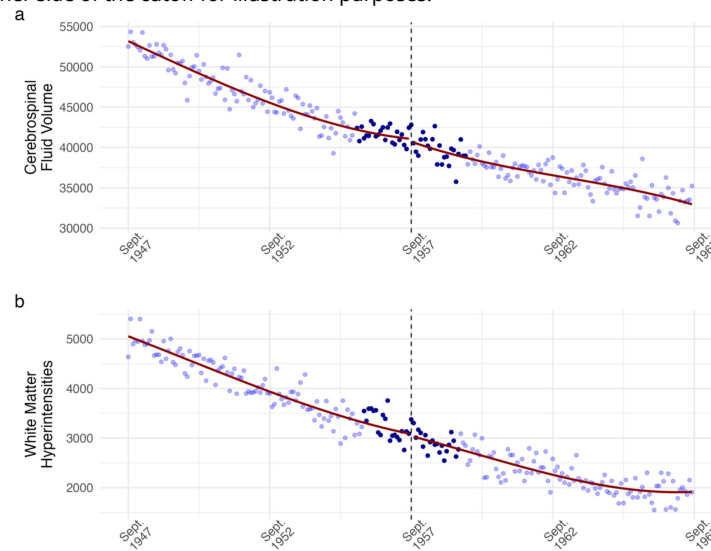
Supplementary Figure 2

This regression discontinuity plot shows the effect of the law on the percentage of students staying an additional year in school (Y). This is the first stage in the two-stage fuzzy regression discontinuity analysis, this increased attendance by roughly 10%. See "first.stage" in Supp. Table 2 for coefficient estimates.



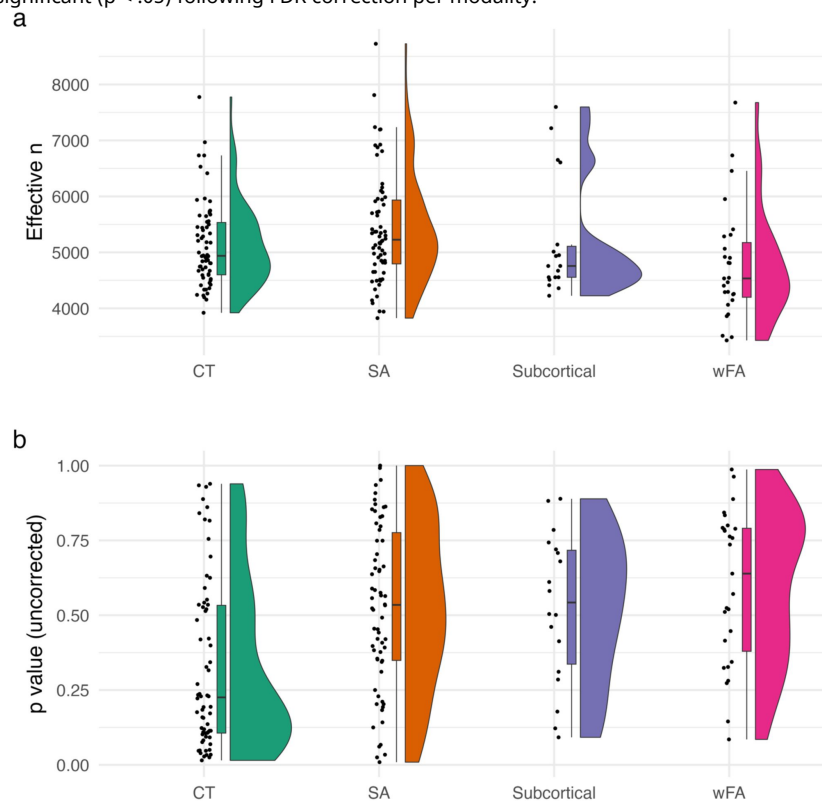
Supplementary Figure 3

Regression discontinuity plot of monthly averaged *a)* cerebrospinal fluid volume and *b)* white matter hyperintensities plotted by the participant's date of birth in months (X; our running variable). The dashed line corresponds to Sept. 1957 when ROSLA came into effect. Dark blue dots represent the mean-squared error-optimized bandwidths, while 3rd order polynomials (in dark red) were fit on either side of the cutoff for illustration purposes.



Supplementary Figure 4

Raincloud plots of *a)* the effective number of observations and *b)* uncorrected p-value of a local-linear fuzzy RD per region per modality [cortical thickness (CT), Surface Area (SA), Subcortical regions, and weighted mean fractional anisotropy (wFA)]. No regions were significant ($p < .05$) following FDR correction per modality.



Supplementary Table 1

Fuzzy RD Placebo Outcome results

[Summer is a dummy coding variable for if the participants date of birth was in July or August. This makes it inherently weighed, *by design*, against ROSLA (which happened on September 1st). These children can be seen in **Sup. Fig 2** [↗](#), see methods for further details].

Fuzzy RD Parameter	bandwidth	eff.obs	estimate	Confidence Interval	p.value	pFDR
<i>sex ~ EduAge16 running_var</i>	23.41	8008.29	-0.28	(-0.76, 0.20)	0.266	0.384
<i>t2_FLAIR ~ EduAge16 running_var</i>	23.41	5112.74	-0.28	(-0.49, -0.07)	0.009	0.059
<i>visit_day_correct ~ EduAge16 running_var</i>	27.44	9365.44	1023.63	(123.00, 1924.25)	0.025	0.083
<i>visit_day_correct² ~ EduAge16 running_var</i>	25.61	8733.81	3675359.31	(52771.30, 7297947.32)	0.047	0.101
<i>headmotion ~ EduAge16 running_var</i>	27.05	6088.77	0.09	(-0.10, 0.28)	0.343	0.406
<i>summer ~ EduAge16 running_var</i>	20.14	6904.60	-4.40	(-5.64, -3.17)	0.000	0.000 ²
<i>imaging_center_11025 ~ EduAge16 running_var</i>	21.74	7380.04	-0.56	(-1.06, -0.07)	0.025	0.083
<i>imaging_center_11026 ~ EduAge16 running_var</i>	25.18	8496.55	0.18	(-0.15, 0.51)	0.296	0.385
<i>imaging_center_11027 ~ EduAge16 running_var</i>	25.27	8527.40	0.27	(-0.15, 0.70)	0.215	0.350
<i>imaging_center_11028 ~ EduAge16 running_var</i>	23.11	7818.34	0.17	(-0.06, 0.40)	0.141	0.262
<i>dMRI_25922_1 ~ EduAge16 running_var</i>	27.06	5591.14	0.09	(-0.50, 0.67)	0.778	0.843
<i>dMRI_25921_1 ~ EduAge16 running_var</i>	27.03	5773.50	0.02	(-0.54, 0.57)	0.953	0.953
<i>dMRI_25928_1 ~ EduAge16 running_var</i>	26.28	5454.88	0.59	(0.01, 1.16)	0.046	0.101

Supplementary Table 2

Fuzzy RD Global Neuroimaging Results

Fuzzy RD Parameter	eff.obs	bandwidth	estimate (Y)	Confidence Interval	p.value	first.stage
<i>Surface Area</i>	7305.947	35.438	1055.733	(-13654.9, 15766.37)	0.889	0.084
<i>Cortical Thickness</i>	4606.983	22.242	-0.107	(-0.235, 0.02)	0.099	0.099
<i>White Matter Hyperintensities</i>	4238.597	20.637	1440.618	(-1835.03, 4716.27)	0.395	0.101
<i>CSF normalized for head size</i>	4857.130	23.479	2989.622	(-15306.03, 21285.27)	0.753	0.097
<i>TBV normalized for head size</i>	5373.601	26.231	4602.620	(-70961.58, 80166.82)	0.906	0.095
<i>Mean Weighted FA</i>	4362.964	21.238	-0.007	(-0.02, 0.01)	0.466	0.100

Supplementary Table 3

Fuzzy RD Uncorrected Global Neuroimaging Results

<i>Fuzzy RD Parameter Uncorrected</i>	eff.obs	bandwidth	estimate (uY)	Confidence Interval	p.value	first.stage
<i>Uncorrected Surface Area</i>	7939.988	38.722	-5267.958	(-23483.93, 12948.01)	0.577	0.082
<i>Uncorrected Cortical Thickness</i>	4749.844	23.000	-0.139	(-0.275, -0.005)	0.043	0.098
<i>Uncorrected White Matter Hyperintensities</i>	4467.847	21.778	2668.595	(-856.54, 6193.73)	0.141	0.099
<i>Uncorrected CSF normalized for head size</i>	4930.766	23.867	2968.307	(-16016.26, 21952.87)	0.764	0.097
<i>Uncorrected TBV normalized for head size</i>	5462.770	26.634	-17635.849	(-96743.57, 61471.87)	0.667	0.094
<i>Uncorrected Mean Weighted FA</i>	4422.511	21.512	-0.008	(-0.03, 0.01)	0.404	0.100

Supplementary Table 4

One month window Bayesian Analysis

Only participants born in August and September 1957 are included. Bayesian analysis of global neuroimaging measures (Y) using a local randomization RD ("ROSLA"; dummy coding for participants born in September 1957) and the associational effect (EduAge) of the amount of attained education in years. The estimate is the median of the posterior reported in raw units. Std. Est is the standardized estimate (for EduAge this is a standardized continuous variable, while ROSLA is a dummy coded variable with 1 corresponding to being impacted by the policy). The estimate, CI & BF are reported for a normal prior (mean = 0, SD = 1). We also report Bayes factors (BF) for other priors. Two measures CSF & TBV are normalized for head size, this is reflected with "_norm".

Y	X	n	Estimate	Std Est.	95 Credible Interval	BF	<i>normal(0, 0.5)</i>	<i>normal(0, 1.5)</i>
SA	ROSLA	232	5.72	0.00	(-3144.86, 3131.35)	BF ₀₁ =18.21	BF ₀₁ =9.22	BF ₀₁ =27.25
SA	EduAge	229	833.24	0.14	(248.86, 1438.26)	BF ₁₀ =2.31	BF ₁₀ =4.71	BF ₁₀ =1.59
CT	ROSLA	232	0.00	-0.02	(-0.03, 0.02)	BF ₀₁ =15.09	BF ₀₁ =7.61	BF ₀₁ =22.48
CT	EduAge	229	0.00	-0.02	(-0.01, 0.01)	BF ₀₁ =14.38	BF ₀₁ =7.36	BF ₀₁ =21.33
CSF_norm	ROSLA	232	-664.54	-0.04	(-4828.25, 3446.43)	BF ₀₁ =14.50	BF ₀₁ =7.34	BF ₀₁ =21.65
CSF_norm	EduAge	229	775.82	0.12	(-11.4, 1558.66)	BF ₀₁ =2.33	BF ₀₁ =1.14	BF ₀₁ =3.58
TBV_norm	ROSLA	232	3884.91	0.06	(-12228.83, 20281.2)	BF ₀₁ =13.61	BF ₀₁ =6.90	BF ₀₁ =20.30
TBV_norm	EduAge	229	-405.08	-0.02	(-3491.44, 2704.24)	BF ₀₁ =14.47	BF ₀₁ =7.22	BF ₀₁ =21.74
WM_hyper	ROSLA	232	222.81	0.07	(-649.46, 1109.96)	BF ₀₁ =13.26	BF ₀₁ =6.78	BF ₀₁ =20.04
WM_hyper	EduAge	229	-56.48	-0.04	(-224.85, 114.12)	BF ₀₁ =12.04	BF ₀₁ =6.01	BF ₀₁ =17.93
wFA	ROSLA	228	0.00	-0.09	(0, 0)	BF ₀₁ =11.63	BF ₀₁ =6.00	BF ₀₁ =17.29
wFA	EduAge	228	0.00	0.04	(0, 0)	BF ₀₁ =11.47	BF ₀₁ =5.78	BF ₀₁ =17.26

Supplementary Table 5

One Month window Test of Covariates

A Bayesian local randomization analysis of placebo outcomes with participants born in August and September 1957 included. ROSLA dummy codes participants born in September 1957. Placebo outcomes are a common method to falsify an RD design, as seen above, by definition they should be unrelated to the natural experiment (ROSLA). The estimate is the median of the posterior. The estimate, CI & BF are reported for a normal prior (mean = 0, SD = 1).

	Y	X	n	Estimate	Bayes Factor	ci_low	ci_high
	sex	ROSLA	394	-0.04	BF ₀₁ =13.45	-0.14	0.05
	visit_day_correct	ROSLA	394	91.82	BF ₀₁ =12.17	-91.10	274.77
	visit_day_correct ²	ROSLA	394	320641.85	BF ₀₁ =14.03	-429953.44	1075987.24
	headmotion	ROSLA	260	0.01	BF ₀₁ =15.15	-0.03	0.05
	imaging_center_11026	ROSLA	391	0.03	BF ₀₁ =14.51	-0.04	0.10
	imaging_center_11027	ROSLA	391	0.04	BF ₀₁ =12.66	-0.05	0.13
	dMRI_25922_1	ROSLA	232	0.02	BF ₀₁ =13.91	-0.09	0.14
	dMRI_25921_1	ROSLA	239	-0.12	BF ₀₁ =2.25	-0.24	0.00
	dMRI_25928_1	ROSLA	233	0.05	BF ₀₁ =10.44	-0.06	0.16
	imaging_center_11025	ROSLA	391	-0.10	BF ₀₁ =2.42	-0.20	0.00

Supplementary Table 6

Five Month window Bayesian Analysis

Bayesian analysis of global neuroimaging measures (Y) using a local randomization RD ("ROSLA"). ROSLA using dummy coding TRUE for participants born from September 1957 until Jan 1958 and FALSE for participants born April-August 1957). EduAge is the associational effect of the amount of attained education in years of the same group of participants. The estimate is the median of the posterior reported in raw units. Std. Est is the standardized estimate (for EduAge this is a standardized continuous variable, while ROSLA is a dummy coded variable with 1 corresponding to being impacted by the policy). The estimate, CI & BF are reported for a normal prior (mean = 0, SD = 1). We also report Bayes factors (BF) for other priors. The suffix "_norm" means normalized for head size.

	Y	X	n	Estimate	Std Est.	95 Credible Interval	BF	normal(0, 0.5)	normal(0, 1.5)
	SA	ROSLA	1195	901.66	0.06	(-491.02, 2329.98)	BF ₀₁ =9.17	BF ₀₁ =9.61	BF ₀₁ =28.59
	SA	EduAge	1185	594.31	0.10	(305.26, 880.86)	BF ₁₀ =41.70	BF ₁₀ =124.31	BF ₁₀ =55.39
	CT	ROSLA	1195	-0.02	-0.10	(-0.02, 0)	BF ₀₁ =7.22	BF ₀₁ =3.58	BF ₀₁ =11.00
	CT	EduAge	1185	0.00	-0.05	(-0.01, 0)	BF ₀₁ =8.81	BF ₀₁ =4.48	BF ₀₁ =13.13
	CSF_norm	ROSLA	1203	-788.23	-0.05	(-2524.2, 971.99)	BF ₀₁ =24.18	BF ₀₁ =11.91	BF ₀₁ =36.13
	CSF_norm	EduAge	1193	761.90	0.12	(409.83, 1123.57)	BF ₁₀ =80.70	BF ₁₀ =381.73	BF ₁₀ =65.54
	TBV_norm	ROSLA	1203	3670.20	0.06	(-3146.49, 10523.16)	BF ₀₁ =20.28	BF ₀₁ =10.22	BF ₀₁ =30.81
	TBV_norm	EduAge	1193	-1476.60	-0.06	(-2881.9, -68.34)	BF ₀₁ =4.21	BF ₀₁ =2.11	BF ₀₁ =6.22
	WM_hyper	ROSLA	1193	2.07	0.00	(-492.03, 501.21)	BF ₀₁ =34.68	BF ₀₁ =17.62	BF ₀₁ =52.25
	WM_hyper	EduAge	1183	38.22	0.02	(-64.51, 137.28)	BF ₀₁ =26.16	BF ₀₁ =13.31	BF ₀₁ =39.37
	wFA	ROSLA	1184	0.00	-0.02	(0, 0)	BF ₀₁ =32.68	BF ₀₁ =16.40	BF ₀₁ =48.70
	wFA	EduAge	1184	0.00	0.06	(0, 0)	BF ₀₁ =3.69	BF ₀₁ =1.87	BF ₀₁ =5.51

Y	X	n	Estimate	Bayes Factor	ci_low	ci_high
<i>sex</i>	ROSLA	1972	-0.04	BF ₀₁ =12.30	-0.08	0.01
<i>visit_day_correct</i>	ROSLA	1972	107.29	BF ₀₁ =1.78	25.87	191.74
<i>visit_day_correct</i> ²	ROSLA	1972	384967.33	BF ₀₁ =3.35	57251.82	720507.66
<i>headmotion</i>	ROSLA	1305	0.01	BF ₀₁ =10.67	0.00	0.03
<i>imaging_center_11026</i>	ROSLA	1950	0.01	BF ₀₁ =34.58	-0.02	0.04
<i>imaging_center_11027</i>	ROSLA	1950	0.02	BF ₀₁ =28.85	-0.02	0.06
<i>dMRI_25922_1</i>	ROSLA	1203	-0.01	BF ₀₁ =33.09	-0.06	0.04
<i>dMRI_25921_1</i>	ROSLA	1246	0.03	BF ₀₁ =19.13	-0.02	0.08
<i>dMRI_25928_1</i>	ROSLA	1208	0.07	BF ₀₁ =1.10	0.02	0.11
<i>imaging_center_11028</i>	ROSLA	1950	0.02	BF ₀₁ =15.85	-0.01	0.04
<i>imaging_center_11025</i>	ROSLA	1950	-0.04	BF ₀₁ =6.20	-0.09	0.00

Supplementary Table 7

Five Month window Test of Covariates

Five-month bandwidth Bayesian local randomization analysis of placebo outcomes. Placebo outcomes are a common method to falsify an RD design, by definition, they should be unrelated to the natural experiment (ROSLA). ROSLA dummy codes participants born after September 1st, 1957. The estimate is the median of the posterior. The estimate, CI & BF are reported for a normal prior (mean = 0, SD = 1).

References

1. (1950) **Geneva Convention IV**
2. (1989) **United Nations Convention on the Rights of the Child**
3. Lövdén M., Fratiglioni L., Glymour M. M., Lindenberg U., Tucker-Drob E. M (2020) **Education and Cognitive Functioning Across the Life Span** *Psychol. Sci. Public Interest* **21**:6–41
4. Strand B. H., et al. (2010) **Educational inequalities in mortality over four decades in Norway: prospective study of middle aged men and women followed for cause specific mortality, 1960–2000** *BMJ* **340**
5. Ross C. E., Wu C. L (1996) **Education, age, and the cumulative advantage in health** *J. Health Soc. Behav* **37**:104–120
6. Kobayashi L. C., et al. (2017) **Childhood deprivation and later-life cognitive function in a population-based study of older rural South Africans** *Soc. Sci. Med* **190**:20–28
7. Waldstein S. R., et al. (2017) **Differential Associations of Socioeconomic Status With Global Brain Volumes and White Matter Lesions in African American and White Adults: the HANDLS SCAN Study** *Psychosom. Med* **79**:327–335
8. Livingston G., et al. (2020) **Dementia prevention, intervention, and care: 2020 report of the Lancet Commission** *Lancet* **396**:413–446
9. Montez J. K., Hummer R. A., Hayward M. D (2012) **Educational attainment and adult mortality in the United States: a systematic analysis of functional form** *Demography* **49**:315–336
10. Turrini S., et al. (2023) **The multifactorial nature of healthy brain ageing: Brain changes, functional decline and protective factors** *Ageing Res. Rev* **88**
11. Ceci S. J (1991) **How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence** *Dev. Psychol* **27**:703–722
12. Ritchie S. J., Tucker-Drob E. M (2018) **How Much Does Education Improve Intelligence? A Meta-Analysis** *Psychol. Sci* **29**:1358–1369
13. Means M. L., Voss J. F (1996) **Who Reasons Well? Two Studies of Informal Reasoning Among Children of Different Grade, Ability, and Knowledge Levels** *Cogn. Instr* **14**:139–178
14. Judd N., Sauce B., Klingberg T (2022) **Schooling substantially improves intelligence, but neither lessens nor widens the impacts of socioeconomics and genetics** *NPJ Sci Learn* **7**
15. Cabeza R., et al. (2018) **Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing** *Nat. Rev. Neurosci* **19**:701–710
16. Buckner R. L (2004) **Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate** *Neuron* **44**:195–208

17. Stern Y (2002) **What is cognitive reserve? Theory and research application of the reserve concept** *J. Int. Neuropsychol. Soc* **8**:448–460
18. Lindenberger U., Lövdén M (2019) **Brain Plasticity in Human Lifespan Development: The Exploration–Selection–Refinement Model** *Annu. Rev. Dev. Psychol* **1**:197–222
19. Newcombe N. S (2011) **What is neoconstructivism?** *Child Dev. Perspect* **5**:157–160
20. Walhovd K. B., et al. (2022) **Education and Income Show Heterogeneous Relationships to Lifespan Brain and Cognitive Differences Across European and US Cohorts** *Cereb. Cortex* **32**:839–854
21. Noble K. G., et al. (2012) **Hippocampal volume varies with educational attainment across the life-span** *Front. Hum. Neurosci* **6**
22. Chan M. Y., et al. (2018) **Socioeconomic status moderates age-related differences in the brain’s functional network organization and anatomy across the adult lifespan** *Proceedings of the National Academy of Sciences* **115**:E5144–E5153
23. Kim J. P., et al. (2015) **Effects of education on aging-related cortical thinning among cognitively normal individuals** *Neurology* **85**:806–812
24. Liu Y., et al. (2012) **Education increases reserve against Alzheimer’s disease—evidence from structural MRI analysis** *Neuroradiology* **54**:929–938
25. Hart S. A., Little C., van Bergen E (2021) **Nurture might be nature: cautionary tales and proposed solutions** *NPJ Sci Learn* **6**
26. Jung R. E., Haier R. J (2007) **The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence** *Behav. Brain Sci* **30**:135–54
27. Farah M. J (2017) **The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences** *Neuron* **96**:56–71
28. Bottenhorn K. L., Sukumaran K., Habre R., Schwartz J., Herting M. (2023) **Air pollution from biomass burning disrupts early adolescent cortical microarchitecture development** *bioRxiv* <https://doi.org/10.1101/2023.10.21.563430>
29. Branigan A. R., McCallum K. J., Freese J (2013) **Variation in the Heritability of Educational Attainment: An International Meta-Analysis** *Soc. Forces* **92**:109–140
30. Lee J. J., et al. (2018) **Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals** *Nat. Genet* **50**:1112–1121
31. Lee D. S., Lemieux T (2010) **Regression Discontinuity Designs in Economics** *J. Econ. Lit* **48**:281–355
32. Cattaneo M. D., Titiunik R. (2022) **Regression discontinuity designs** *Annu. Rev. Econom.*
33. Angrist J. D., Krueger A. B (1991) **Does Compulsory School Attendance Affect Schooling and Earnings?** *Q. J. Econ* **106**:979–1014

34. Armstrong T. B., Kolesár M (2018) **OPTIMAL INFERENCE IN A CLASS OF REGRESSION MODELS** *Econometrica* **86**:655–683
35. Armstrong T. B., Kolesár M (2020) **Simple and honest confidence intervals in nonparametric regression** *Quant. Econom* **11**:1–39
36. Cattaneo M. D., Idrobo N., Titiunik R (2019) **A practical introduction to regression discontinuity designs: Foundations** *arXiv*
37. (1972) **The Raising of the School Leaving Age Order**
38. Davies N. M., Dickson M., Davey Smith G., van den Berg G. J., Windmeijer F. (2018) **The Causal Effects of Education on Health Outcomes in the UK Biobank** *Nat Hum Behav* **2**:117–125
39. Gorman E (2023) **Does Schooling Have Lasting Effects on Cognitive Function? Evidence From Compulsory Schooling Laws** *Demography* **60**:1139–1161
40. Clark D., Royer H (2013) **The Effect of Education on Adult Health and Mortality: Evidence from Britain** *Am. Econ. Rev* **103**:2087–2120 <https://doi.org/10.3386/w16013>
41. Clark D., Royer H (2013) **The Effect of Education on Adult Mortality and Health: Evidence from Britain** *Am. Econ. Rev* **103**:2087–2120 <https://doi.org/10.3386/w16013>
42. Barcellos S. H., Carvalho L. S., Turley P (2023) **DISTRIBUTIONAL EFFECTS OF EDUCATION ON HEALTH** *J. Hum. Resour* **58**:1273–1306
43. Barcellos S. H., Carvalho L. S., Turley P (2018) **Education can reduce health differences related to genetic risk of obesity** *Proc. Natl. Acad. Sci. U. S. A* **115**:E9765–E9772
44. Alfaro-Almagro F., et al. (2018) **Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank** *Neuroimage* **166**:400–424
45. Geruso M., Royer H (2018) **The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK NBER** <https://doi.org/10.3386/w24332>
46. Cattaneo M. D., Jansson M., Ma X (2018) **Manipulation Testing Based on Density Discontinuity** *Stata J* **18**:234–261
47. Desikan R. S., et al. (2006) **An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest** *Neuroimage* **31**:968–980
48. de Groot M., et al. (2013) **Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration** *Neuroimage* **76**:400–411
49. Cattaneo M. D., Idrobo N., Titiunik R (2023) **A Practical Introduction to Regression Discontinuity Designs: Extensions** *arXiv*
50. Marek S., et al. (2022) **Reproducible brain-wide association studies require thousands of individuals** *Nature* **603**:654–660
51. Jeffreys H. (1961) **Theory of probability** Oxford University Press
52. Draganski B., et al. (2006) **Temporal and spatial dynamics of brain structure changes during extensive learning** *J. Neurosci* **26**:6314–6317

53. Draganski B., et al. (2004) **Neuroplasticity: changes in grey matter induced by training** *Nature* **427**:311–312
54. Bråthen A. C. S., et al. (2022) **Cognitive and hippocampal changes weeks and years after memory training** *Sci. Rep* **12**
55. Wenger E., Brozzoli C., Lindenberger U., Lövdén M (2017) **Expansion and Renormalization of Human Brain Structure During Skill Acquisition** *Trends Cogn. Sci* **21**:930–939
56. Garzón B., et al. (2023) **Cortical changes during the learning of sequences of simultaneous finger presses** *Imaging Neuroscience* **1**:1–26
57. Mediavilla T., et al. (2022) **Learning-related contraction of gray matter in rodent sensorimotor cortex is associated with adaptive myelination** *Elife* **11**
58. Schmidt S., et al. (2021) **Experience-dependent structural plasticity in the adult brain: How the learning brain grows** *Neuroimage* **225**
59. Goriounova N. A., et al. (2018) **Large and fast human pyramidal neurons associate with intelligence** *Elife* **7**
60. Vachha B., Huang S. Y (2021) **MRI with ultrahigh field strength and high-performance gradients: challenges and opportunities for clinical neuroimaging at 7 T and beyond** *Eur Radiol Exp* **5**
61. Perneczky R., et al. (2009) **Education attenuates the effect of medial temporal lobe atrophy on cognitive function in Alzheimer’s disease: the MIRAGE study** *J. Alzheimers. Dis* **17**:855–862
62. López-Otin C., Blasco M. A., Partridge L., Serrano M., Kroemer G (2023) **Hallmarks of aging: An expanding universe** *Cell* **186**:243–278
63. Nyberg L., et al. (2021) **Educational attainment does not influence brain aging** *Proc. Natl. Acad. Sci. U. S. A* **118**
64. Burt S. A., Johnson W (2023) **Joint Consideration of Means and Variances Might Change the Understanding of Etiology** *Perspect. Psychol. Sci* **18**:416–427
65. von Stumm S., d’Apice K. (2021) **From Genome-Wide to Environment-Wide: Capturing the Environome** *Perspect. Psychol. Sci* **17**:30–40
66. Pfeifer J. H., Allen N. B., Byrne M. L., Mills K. L (2018) **Modeling Developmental Change: Contemporary Approaches to Key Methodological Challenges in Developmental Neuroimaging** *Dev. Cogn. Neurosci* **33**:1–4
67. Molenaar P. C. M (2004) **A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever** *Measurement* **2**:201–218
68. Schwartz G. L., Maria Glymour M (2023) **Bridging the Divide: Tackling Tensions Between Life-Course Epidemiology and Causal Inference** *Annual Review of Developmental Psychology* **5**:355–374

69. Dolan C. V., Huijskens R. C. A., Minică C. C., Neale M. C., Boomsma D. I (2021) **Incorporating Polygenic Risk Scores in the ACE Twin Model to Estimate A-C Covariance** *Behav. Genet* **51**:237–249
70. Littlejohns T. J., et al. (2020) **The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions** *Nat. Commun* **11**
71. Fischl B., et al. (2002) **Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain** *Neuron* **33**:341–355
72. McCarthy P. (2020) **Funpack** Zenodo <https://doi.org/10.5281/zenodo.3761702>
73. Calonico S., Cattaneo M. D., Titiunik R (2014) **Robust nonparametric confidence intervals for regression-discontinuity designs** *Econometrica* **82**:2295–2326
74. van Buuren S., Groothuis-Oudshoorn K. (2011) **mice: Multivariate Imputation by Chained Equations in R** *Journal of Statistical Software, Articles* **45**:1–67
75. Goodrich B., Gabry J., Ali I., Brilleman S. (2024) **rstanarm: Bayesian applied regression modeling via Stan R package**
76. Depaoli S., van de Schoot R (2017) **Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist** *Psychol. Methods* **22**:240–261
77. Makowski D., Ben-Shachar M. S (2019) **bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework** *Journal of Open Source*
78. Ritchie S. J., et al. (2018) **Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants** *Cereb. Cortex* **28**:2959–2975
79. Avendano M., de Coulon A., Nafilyan V (2020) **Does longer compulsory schooling affect mental health? Evidence from a British reform** *J. Public Econ* **183**
80. Smith Stephen, Almagro Fidel Alfaro, Miller Karla (2024) **UK Biobank Brain Imaging Documentation**

Editors

Reviewing Editor

Jason Lerch

University of Oxford, Oxford, United Kingdom

Senior Editor

Timothy Behrens

University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public review):

Summary:

This fascinating manuscript studies the effect of education on brain structure through a natural experiment. Leveraging the UK BioBank, these authors study the causal effect of

education using causal inference methodology that focuses on legislation for an additional mandatory year of education in a regression discontinuity design.

Strengths:

The methodological novelty and study design were viewed as strong, as was the import of the question under study. The evidence presented is solid. The work will be of broad interest to neuroscientists

Weaknesses:

There were several areas which might be strengthened from additional consideration from a methodological perspective.

<https://doi.org/10.7554/eLife.101526.1.sa3>

Reviewer #2 (Public review):

Summary:

The authors conduct a causal analysis of years of secondary education on brain structure in late life. They use a regression discontinuity analysis to measure the impact of a UK law change in 1972 that increased the years of mandatory education by 1 year. Using brain imaging data from the UK Biobank, they find essentially no evidence for 1 additional year of education altering brain structure in adulthood.

Strengths:

The authors pre-registered the study and the regression discontinuity was very carefully described and conducted. They completed a large number of diagnostic and alternate analyses to allow for different possible features in the data. (Unlike a positive finding, a negative finding is only bolstered by additional alternative analyses).

Weaknesses:

While the work is of high quality for the precise question asked, ultimately the exposure (1 additional year of education) is a very modest manipulation and the outcome is measured long after the intervention. Thus a null finding here is completely consistent educational attainment (EA) in fact having an impact on brain structure, where EA may reflect elements of training after a second education (e.g. university, post-graduate qualifications, etc) and not just stopping education at 16 yrs yes/no.

The work also does not address the impact of the UK Biobank's well-known healthy volunteer bias (Fry et al., 2017) which is yet further magnified in the imaging extension study (Littlejohns et al., 2020). Under-representation of people with low EA will dilute the effects of EA and impact the interpretation of these results.

References:

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9), 1026-1034. <https://doi.org/10.1093/aje/kwx246>

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boulton, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., ... Allen, N. E. (2020). The UK

Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications*, 11(1), 2624. <https://doi.org/10.1038/s41467-020-15948-9>

<https://doi.org/10.7554/eLife.101526.1.sa2>

Reviewer #3 (Public review):

Summary:

This study investigates evidence for a hypothesised, causal relationship between education, specifically the number of years spent in school, and brain structure as measured by common brain phenotypes such as surface area, cortical thickness, total volume, and diffusivity.

To test their hypothesis, the authors rely on a "natural" intervention, that is, the 1972 ROSLA act that mandated an extra year of education for all 15-year-olds. The study's aim is to determine potential discontinuities in the outcomes of interest at the time of the policy change, which would indicate a causal dependence. Naturalistic experiments of this kind are akin to randomised controlled trials, the gold standard for answering questions of causality.

Using two complementary, regression-based approaches, the authors find no discernible effect of spending an extra year in primary education on brain structure. The authors further demonstrate that observational studies showing an effect between education and brain structure may be confounded and thus unreliable when assessing causal relationships.

Strengths:

- (1) A clear strength of this study is the large sample size totalling up to 30k participants from the UK Biobank. Although sample sizes for individual analyses are an order of magnitude smaller, most neuroimaging studies usually have to rely on much smaller samples.
- (2) This study has been preregistered in advance, detailing the authors' scientific question, planned method of inquiry, and intended analyses, with only minor, justifiable changes in the final analysis.
- (3) The analyses look at both global and local brain measures used as outcomes, thereby assessing a diverse range of brain phenotypes that could be implicated in a causal relationship with a person's level of education.
- (4) The authors use multiple methodological approaches, including validation and sensitivity analyses, to investigate the robustness of their findings and, in the case of correlational analysis, highlight differences with related work by others.
- (5) The extensive discussion of findings and how they relate to the existing, somewhat contradictory literature gives a comprehensive overview of the current state of research in this area.

Weaknesses:

- (1) This study investigates a well-posed but necessarily narrow question in a specific setting: 15-year-old British students born around 1957 who also participated in the UKB imaging study roughly 60 years later. Thus conclusions about the existence or absence of any general effect of the number of years of education on the brain's structure are limited to this specific scenario.
- (2) The authors address potential concerns about the validity of modelling assumptions and the sensitivity of the regression discontinuity design approach. However, the possibility of

selection and cohort bias remains and is not discussed clearly in the paper. Other studies (e.g. Davies et al 2018, <https://www.nature.com/articles/s41562-017-0279-y>) have used the same policy intervention to study other health-related outcomes and have established ROSLA as a valid naturalistic experiment. Still, quoting Davies et al. (2018), "This assumes that the participants who reported leaving school at 15 years of age are a representative sample of the sub-population who left at 15 years of age. If this assumption does not hold, for example, if the sampled participants who left school at 15 years of age were healthier than those in the population, then the estimates could underestimate the differences between the groups.". Recent studies (Tyrrell 2021, Pirastu 2021) have shown that UK Biobank participants are on average healthier than the general population. Moreover, the imaging sub-group has an even stronger "healthy" bias (Lyll 2022).

(3) The modelling approach used in this study requires that all covariates of no interest are equal before and after the cut-off, something that is impossible to test. Mentioned only briefly, the inclusion and exclusion of covariates in the model are not discussed in detail. Standard imaging confounds such as head motion and scanning site have been included but other factors (e.g. physical exercise, smoking, socioeconomic status, genetics, alcohol consumption, etc.) may also play a role.

<https://doi.org/10.7554/eLife.101526.1.sa1>

Author response:

To Reviewer #1:

Thank you for your kind words regarding the novelty, study design, and evidence presented. We will clarify our language when describing fuzzy local-linear regression discontinuity analysis. We thank you for this feedback as our goals are to introduce these methods to a neuroscientific audience. Lastly, we will respond and clarify the methodological points, including post-selection inference, bandwidths, and Bayesian analysis in version 2.

To Reviewers #2 and #3:

We thank you both for your constructive feedback, specifically in highlighting 1) the scope of the intervention and 2) the UKB-neuro healthy volunteer bias. In the next manuscript version, we will expand our discussion of plausible reasons for not finding an effect – weighing up the strengths and limitations of our study in 3 aspects; statistical (RD power), design-based (lack of representativeness vs. large sample), and mechanistic (the impact/or lack thereof of one-year of education on neural plasticity decades later). As we believe the approach of natural experiments with RD designs has considerable promise for the field of population cognitive neuroscience beyond this particular study, we will address each of these points within a broader section focused on considerations on how to optimize the insight, power, and inferences gained in future work within and beyond Biobank. Moreover, we will situate our discussion on the magnitude of the educational intervention among a broader discussion of cognitive training versus education, and short - versus long-term effects. We believe revising the manuscript will improve interpretation for the reader and thank you for your in-depth feedback. Lastly, we will provide a point-by-point response in the next version.

<https://doi.org/10.7554/eLife.101526.1.sa0>