eLife

Reviewed Preprint
v2 • April 17, 2025
Revised by authors

Reviewed Preprint
v1 • November 5, 2024

Neuroscience

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, Netherlands

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. While the finding is **convincing** on its own, as the analysis was pre-registered and very carefully conducted, the impact is limited as the manipulated variable only relates to a single additional year of education (remaining in education to 15 vs 16 years of age).

https://doi.org/10.7554/eLife.101526.2.sa3

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 the UK Biobank. 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 is severely lacking [25]. 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, interacting, sociodemographic and individual characteristics [3,26] – many of which have been associated with brain structure (e.g., intelligence [27], parental income [28], neighborhood pollution [29]). Among the plethora of environmental factors, there is also a substantial (~40% heritability) genetic component implicated in educational attainment, further confounding potential effects [30,31]. 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 [32,33]. 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 [34]. 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 [32,33]. Recent advances in the field of econometrics have optimized inference and largely standardized regression discontinuity (RD) analysis [33,35–37]. Similar natural experimental designs 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') [38]. 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 [39–44]. 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 [45]. 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) [35,36]. 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. Compliance with ROSLA was very high (near 100%; **Sup. Figure 2**). However, given the cultural and historical trends leading to an increase in school attendance before ROSLA, most adolescents were continuing with education past 15 years of age before the policy change (Sup Plot. 7b). Prior work has estimated 25 percent of children would have left school a year earlier if not for ROSLA [41]. Using the UK Biobank, we estimate this proportion to be around 10%, as the sample is healthier and of higher SES than the general population (**Sup. Figure 2**; **Sup. Table 2**) [46–48].

Local-linear RD analysis tests the effect of this policy change 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 [33]. 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* around the cutoff.

Using fuzzy local-linear RD we fit a series of regressions to various measures of brain structure 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. Figure 3**). These results did not change when imputing missing (~4%) covariate data.

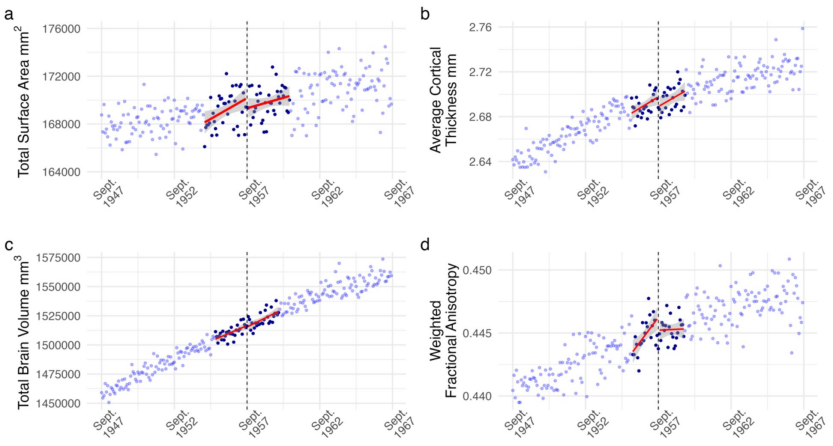


Figure 1

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. Dark blue dots represent all individuals within the mean-squared error-optimized bandwidths, reflecting participants used for analysis with a local-linear fuzzy RD approach. We found no evidence of an effect from an additional year of education on any of the preregistered structural neuroimaging measures across various analysis specifications (**Sup Tables 2**,**3**,**4** & **6**).

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 [33]. 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 [38,49]. However, for completeness we tested this question [50], finding no evidence that individuals were somehow able to adjust their enrollment (T_q = -0.72, p = .47; **Sup. Figure 2**). A second validation approach is to employ *placebo outcome tests* [33]. 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**). Our findings add to the existing body of prior work [39,41,42,44] 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 [51] 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 [52] (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 5**).

Bayesian Local Randomization Robustness Analysis

As an additional preregistered robustness test, we used an alternate framework, 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) [33,53]. 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 (BF_{01} = 18.21), average cortical thickness (BF_{01} = 15.09), total brain volume (BF_{01} = 13.61), weighted fractional anisotropy (BF_{01} = 11.63), white matter hyperintensities (BF_{01} = 13.26), and cerebral spinal fluid volume (BF_{01} = 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** & **6**; **Sup. Figure 6**).

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** & **3**, **Sup. Table 6**). Lastly, we repeated our placebo outcome tests for both a one- and five-month window local randomization analysis, finding no associations (**Sup Table 5** & **7**), demonstrating the robustness of the natural experiment ROSLA and our analysis approach.

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. [54]), 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 ~ 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 (**Figure 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 Jeffrey criteria[55]) in favor of the null when compared to the causal estimate (**Figure 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).

A post hoc replication of this associational analysis in eight additional 10-month cohorts spaced two years apart (**Sup. Figure 7**) indicates our preregistered report on the associational effect of educational attainment on CSF to be most likely a false-positive (**Sup. Figure 8**). Yet, the positive association between surface area and educational attainment is robust across the additional eight replication cohorts.

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: normal(0, 1)] was not preregistered and serves to illustrate our global neural findings.

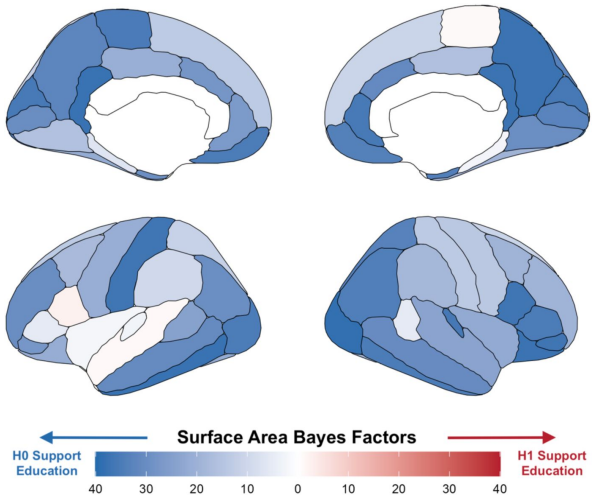
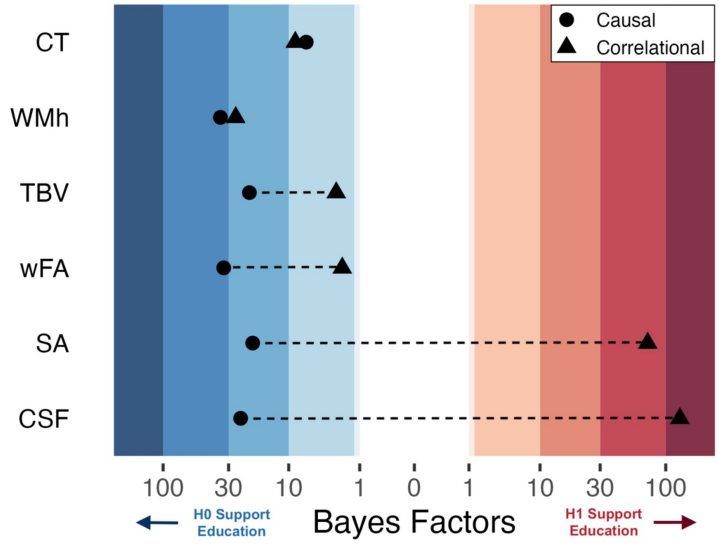


Figure 3

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 ≈ 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. Note: The association between educational attainment and CSF did not hold across eight replication sub-cohorts, yet surface area demonstrated a robust association (**Sup. Figure 7** & **8**).



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 [39,41–44]. 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.

Our robust null result is seemingly at odds with causal inference findings of education's positive behavioral effect on intelligence [11,14,40] – which is sustained throughout decades [12,39]. 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 [56–58]. 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 [42,43]. 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* [59,60], 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 [61]. Possible cellular candidates for initial experience-dependent plasticity are an increase in dendritic spines, the swelling of astrocytes, and intracortical myelin adaptations [61–63]. These structural changes may be detectable using other approaches such as in vivo cellular work (cf. [64]), extreme high field strengths [60,65], or postmortem histology [66,67].

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 [68] – 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 [69] but also equally hard to adequately control[70]. 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 [71–73], as well as more sophisticated models that can capture gene-by-environment-interplay [14,69,74]. Only then will we be able to identify idiosyncratic environmental effects and individual characteristics underlying heterogenous lifespan development.

The UK Biobank is known to have *'healthy volunteer bias'*, as respondents tend to be healthier, more educated, and are more likely to own property [46–48]. Various types of selection bias can occur in non-representative samples, impacting either internal (type 1) or external (type 2) validity [47,75]. One benefit of a natural experimental design is that it protects against threats to internal validity from selection bias [43], design-based internal validity threats still exist, such as if volunteer bias differentially impacts individuals based on the cutoff for assignment. A more pressing limitation – in particular, for an education policy change – is our power to detect effects using a sample of higher-educated individuals. This is evident in our first stage analysis examining the percentage of 15-year-olds impacted by ROSLA, which we estimate to be 10% in neuro-UKB (**Sup. Figure 2** & **Sup. Table 2**), yet has been reported to be 25% in the UK general population [41]. Our results should be interpreted for this subpopulation (UK, 1973, from 15 to 16 years of age, compliers) as we estimate a 'local' average treatment effect [76]. Natural experimental designs such as ours offer the potential for high internal validity at the expense of external validity.

Here, we report a lack of causal evidence of a year of school on long-term neural outcomes in a large sample of aging individuals. 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 [38]. 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**) [41,42]. Crucially for our purposes, ROSLA is a well-studied natural experiment with (commonly agreed, e.g. [39,41–43]) high design validity. A sizable body of prior work has found behavioral effects from the 1972 ROSLA [39,41–44].

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 [77]. We followed our preregistration (*https://osf.io/rv38z*) closely, yet some minor deviations were necessary and explicitly outlined in *'deviations from preregistration'* (code: *https://github.com/njudd/eduBrain*).

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 [51]. 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 [52]. Lastly, we examine subcortical volume in 18 regions [78].

All measures are derived from the image preprocessing pipeline from the UKB, further preprocessing details are outlined elsewhere [45,79]. Outliers were winsorized 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 [32,33] – 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 [49]. Nevertheless, we conducted a density

test of the running variable to check for bunching near the cutoff [50]. Although the design validity of the 1972 ROSLA is well established [39,41–44] we still tested a variety of placebo outcomes (outlined in 'covariates of no interest') – outcomes implausible to be affected by the intervention.

RD designs, like ours, can be 'fuzzy' indicating when assignment only increases the probability of receiving it, in turn, treatment assigned and treatment received do not correspond for some units [33,53]. For instance, due to cultural and historical trends, there was an increase in school attendance before ROSLA; most adolescents were continuing with education past 15 years of age (Sup Plot. 7b). Prior work has estimated that 25 percent of children would have left school a year earlier if not for ROSLA [41]. Using the UK Biobank, we estimate this proportion to be around 10%, as the sample is healthier and of higher SES than the general population (**Sup. Figure 2**; **Sup. Table 2**) [46–48].

RD analysis broadly falls into two separate but complementary frameworks [37,53]. 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 (*w*) 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 *w*. As *w* 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 [32,35,36]. 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 [39]. 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 [35,36].

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 [44] to as small as a year [39]. 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 [37,53]. 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 [80]. 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 [37], 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 [37]. 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 (≈4%) with classification and regression trees from the MICE package [81]. 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 [33]. This alternative framework assumes a small window (w) around the cutoff (c) where the running variable is treated "as if random" [53]. While the local randomization framework invokes stricter assumptions on the assignment mechanism, placing more importance on the window around the cutoff (w = [c − w, c + w]) it handles discrete running variables well [53]. 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 [33,53], we included participants within the smallest window possible (w = 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. As preregistered, in this framework we report the 'intent to treat estimate' (ITT) which should be interpreted as the effect of the policy change ROSLA on neuroimaging outcomes. Lasty, 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 [82]. All priors (p) used a normal distribution centered at 0 with autoscaling [p*sd(y)/sd(x)]. Our preregistration referred to using the *'default'* weakly informative prior of STAN (i.e., 2.5 SDs) [83]. 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 (**Sup. Figure 6**).

Model diagnostics were checked with trace plots, posterior distributions, and rhat values (<1.05) [84]. We then computed Bayes Factors using the bayestestR package [85] using the Savage-Dickey density ratio with a point-null of 0 for each of the 3 priors. Bayes factors are reported as BF_{10} in support of the alternative hypothesis, we report Bayes factors under 1 as the multiplicative inverse (BF_{01} = 1/BF). The strength of evidence was interpreted on a graded scale using the criteria preregistered [55]. 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 (w = 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 (w = 1 & 5 months). Educational attainment is commonly used to report the effect of education on brain structure [86]. The included subjects are identical to the local randomization approach, except the dummy coded term ROSLA was substituted for continuous EduYears.

We hypothesize the *associational* effect of educational attainment on global neuroimaging measures is the result of various unmeasured individual and societal-level factors confounding this relationship. Upon a reviewer-initiated comment (see Elife Reviewer 1 Recommendation 7), we made eight more 10-month sub-cohorts spaced two years apart (4 after ROSLA & before ROSLA; **Sup. Figure 7**). In each of these eight cohorts, we tested the effect of educational attainment on our six global neuroimaging covariates following identical methodology (Bayesian model, normal prior (mean=0, sd=1). From these models, we extracted Bayesian posterior estimates of the associational effect of educational attainment and plotted them (**Sup. Figure 8**).

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 [37]. 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 [87], 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 [41]. This is an artifact of asking in the UK Biobank *'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 [88], 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 [77]. 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 (**Sup. Figure 1**). 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 & DoS^2). 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 [89]. 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 (**Sup. Table 5** & **7**). This was also tested (and confirmed) for the Local-Randomization Framework (w = 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 (uY) 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 ($\beta^{'}_{*}$) and subtracted from the unadjusted outcome (uY) 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. The latest version of the package includes covariate correction and this is preferred to our method outlined below. For the Bayesian local randomization analysis and correlational analysis, we simply included the covariates in the model.

$$uY = \beta_0 + \beta_1(X) + \beta_2(X \geq 0) + \beta_3(X * X \geq 0) + \beta_* \mathbf{C}' + \epsilon$$

$$Y = uY - (\widehat{\beta_*} \times \mathbf{C}')$$

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 falsification 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-weighed 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.

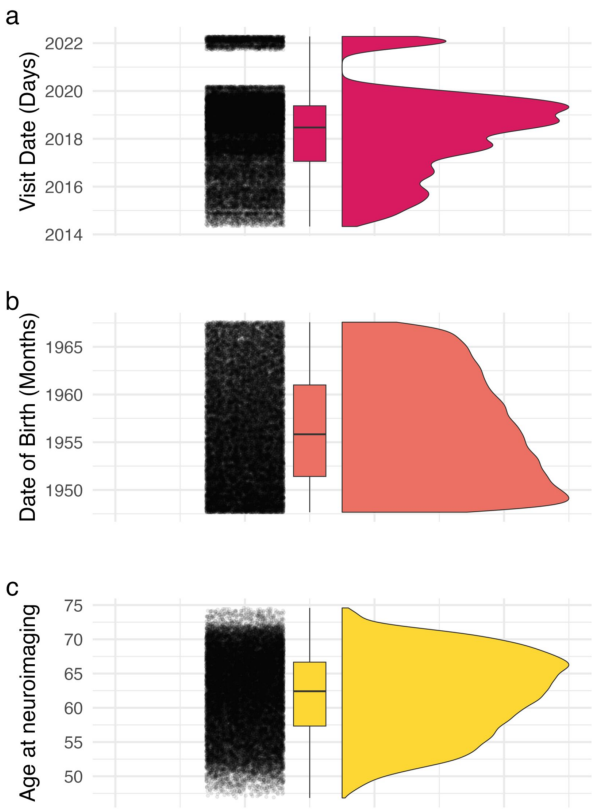
Choice of priors

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 (**Sup. Figure 6**). 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**).

Post hoc associational replication

In our associational analysis, we used educational attainment in years in the identical subset of subjects used for the local randomization causal analysis. This allowed us to compare and contrast the weight of evidence in a causal and associational analysis (**Figure 3**) within the same subjects. Upon a reviewer-initiated comment (see Elife v1 Reviewer 1 Recommendation 7), we made eight additional subcohorts (see **Sup. Figure 7**) spaced two years apart, as the placement of our associational analysis should not matter. We found our results to hold in all global neuroimaging covariates bar CSF, which seemed to only show an effect in our specific preregistered cohort (**Sup. Figure 8**).

Supplementary Information

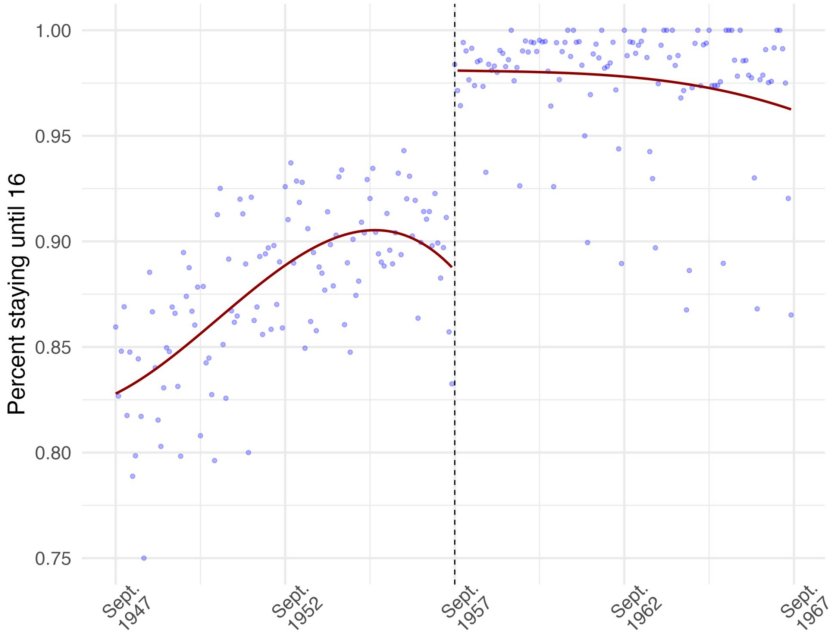


Supplementary Figure 1

Raincloud plots illustrating the distribution of participants ***a)*** date of visit, ***b)*** date of birth, and ***c)*** 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.

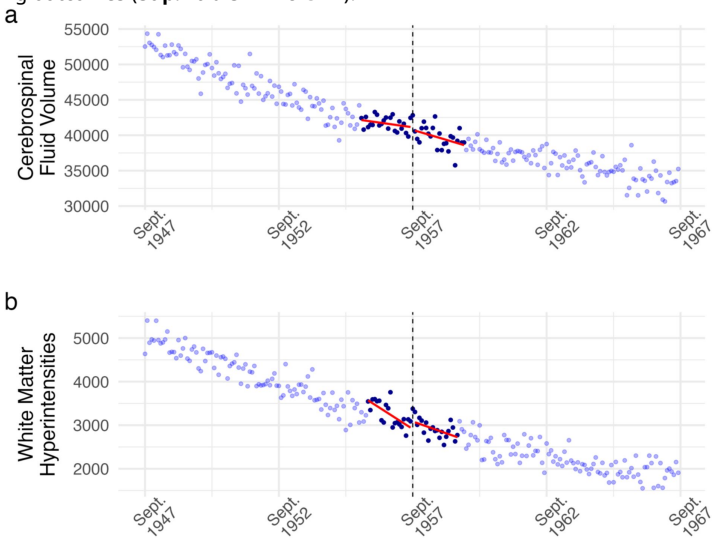
Supplementary Figure 2

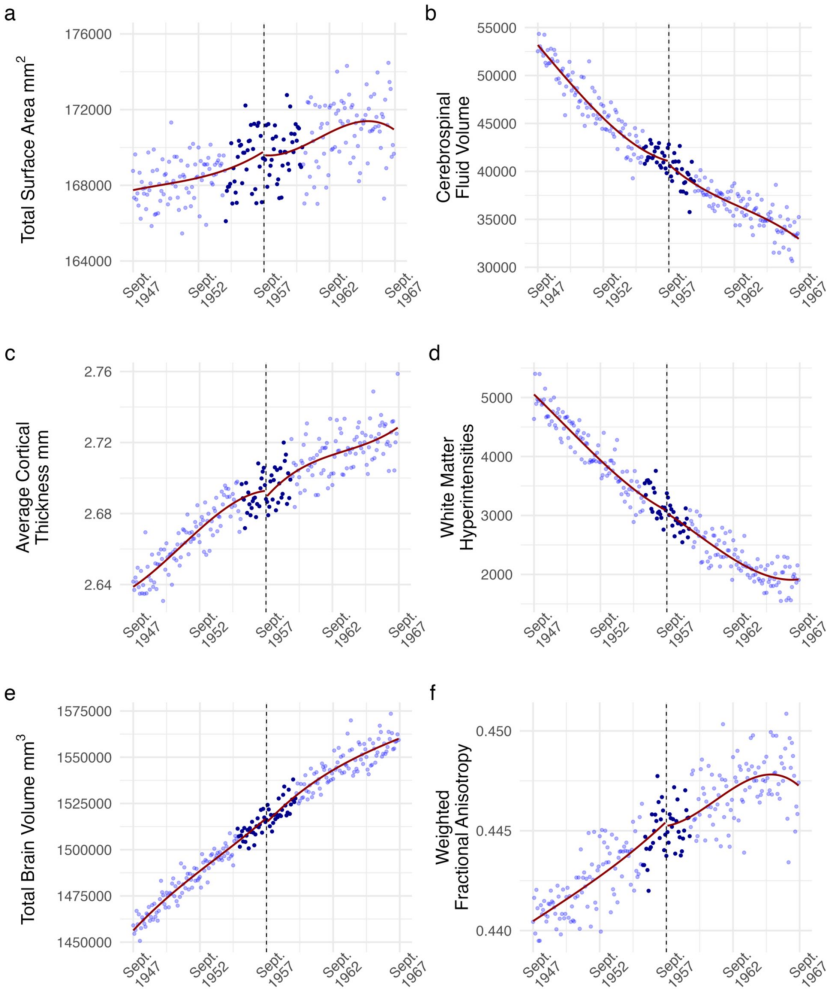
This regression discontinuity plot shows the effect of the law on the percentage of students staying an additional year in school (Y) by date of birth monthly averaged bins. This is the first stage in the two-stage fuzzy regression discontinuity analysis, this increased attendance by roughly 10%. See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



Supplementary Figure 3

Regression discontinuity plots 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. Linear model fits (in red) on either side of the cutoff are added for illustration purposes. We found no effect of an additional year from ROSLA on any of our global neuroimaging outcomes (**Sup. Table 2** & **3**).



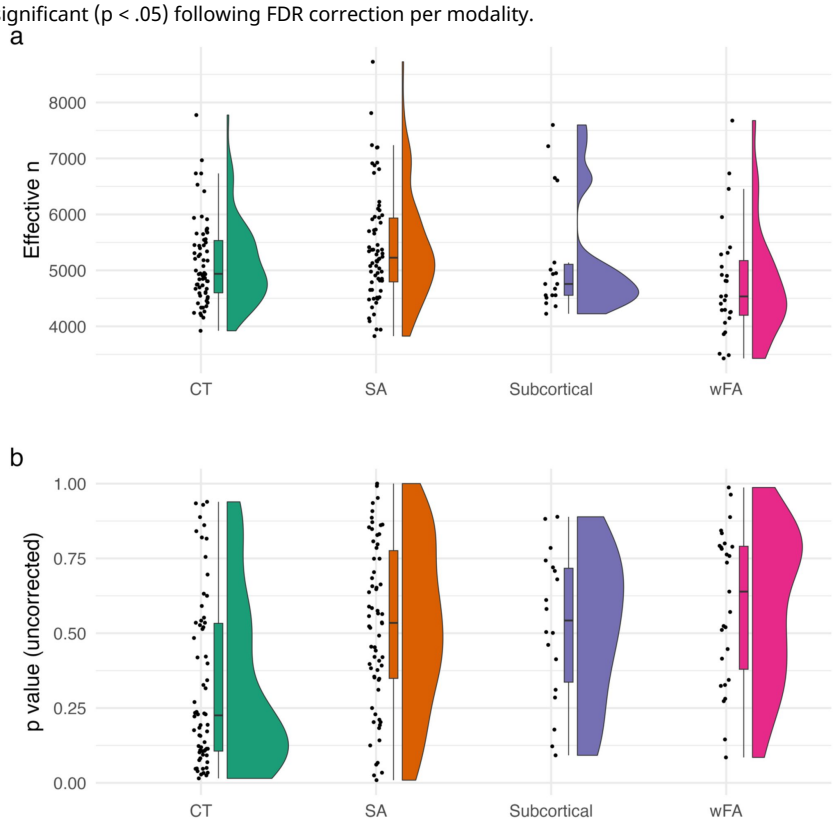


Supplementary Figure 4

Regression discontinuity plots of monthly averaged *a)* total surface area, *b)* cerebrospinal fluid volume, *c)* average cortical thickness, *d)* white matter hyperintensities, *e)* total brain volume, and *f)* weighted fractional anisotropy, plotted by the participant's date of birth in months (X; our running variable). The dashed line corresponds to Sept. 1957 the date of birth inclusion criteria of ROSLA. This plot is the same as **Figure 1** and **Sup. Figure 3**, yet with 3rd order global polynomial fits; a common illustration in the RD literature.

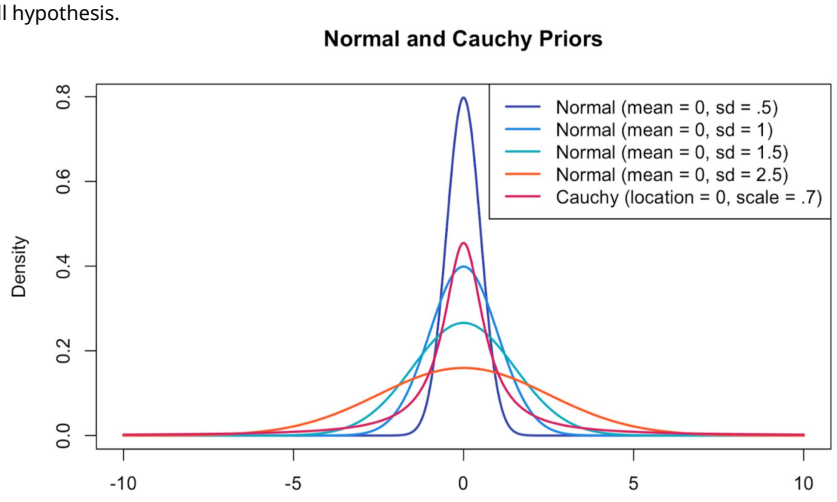
Supplementary Figure 5

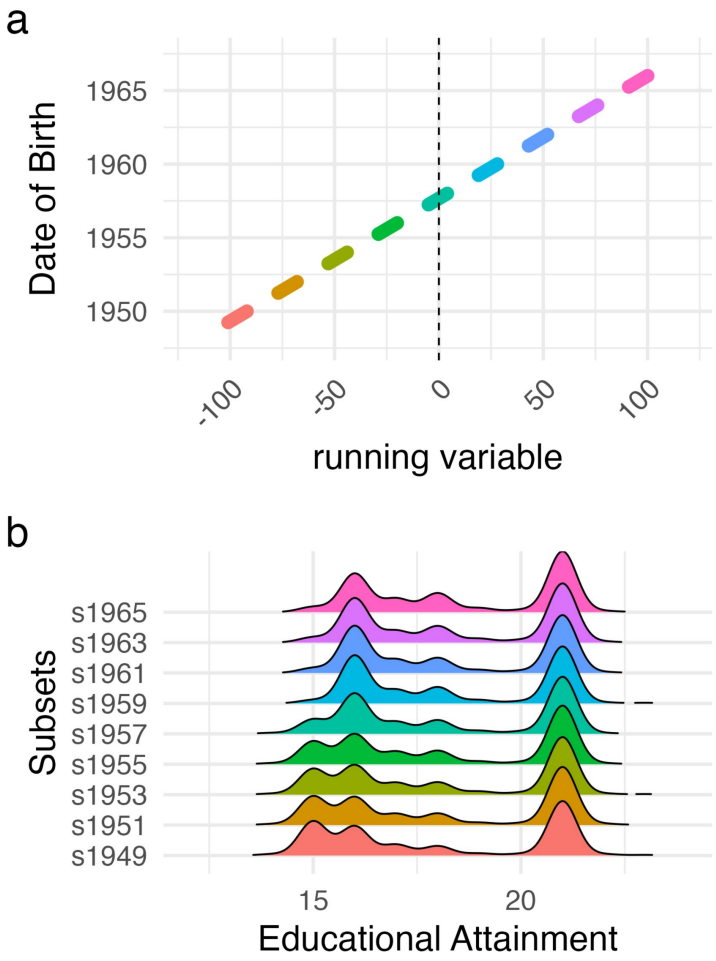
Raincloud plots of ***a)*** the effective number of observations and ***b)*** uncorrected p-value of a local-linear fuzzy RD per region per modality [cortical thickness (CT), Surface Area (SA), Subcortical regions, and weighted mean fractional anisotropy (wFA)]. No regions were significant (p < .05)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



Supplementary Figure 6

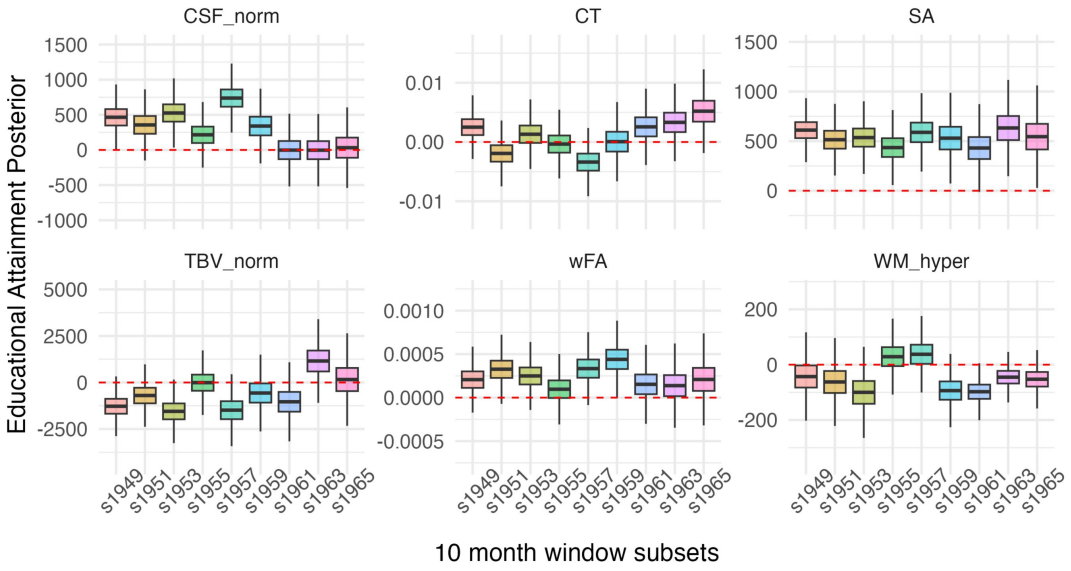
An overview of the Bayesian priors used in our Bayesian local-randomization analysis. All priors have a location of zero yet vary in their scale parameter. The blue priors represent our reported (not preregistered priors). In text, we report the normal distribution with a standard deviation of one. The orange Normal distribution (sd = 2.5) is the default, 'weakly informative' prior from STAN (our initial preregistered prior; which is too wide; thereby offering excessive support for the null hypothesis). The magenta Cauchy distribution is the default 'wide' prior used in the BayesFactor package. Point null Bayes Factors are particularly sensitive to the prior used – yet across all tested specifications (**Sup. Table 4** & **6**) – we find evidence in support of the null hypothesis.





Supplementary Figure 7

An overview of eight cohorts/subsets (plus the original 5-month local randomization cohort; s1957) spaced 2 years apart with 10-month windows of included subjects in **panel a**. The running variable is the date of birth centered around the start of the ROSLA natural experiment (0 = Sept. 1957) indicated by the dashed line. **Panel b** illustrates educational attainment distributions for the nine cohorts; note the lack of individuals with only 15 years of education after 1957 due to ROSLA.



Supplementary Figure 8

Illustrates the posterior of the associational effect of educational attainment on neuroimaging measures across nine 10-month subcohorts of UK Biobank data spaced two years apart. For this posthoc analysis, a normally distributed prior with a mean of zero and an SD of one was used. The cohort 's1957' is our initial preregistered associational analysis, which is centered on the ROSLA cutoff allowing us to compare and contrast a causal parameter to an associational one (**Figure 3**). Crucially, we found evidence in support of an associational effect of education on CSF and Surface Area; along with evidence in support of the null hypothesis for the other global neuroimaging covariates. Upon a reviewer-initiated comment (see Elife v1 Reviewer 1 Recommendation 7), we made additional subcohorts (see **Sup. Figure 7**) as the placement of our associational analysis should not matter. This additional post hoc analysis seems to indicate the effect of educational attainment on CSF to be a false-positive – illustrated by the additional eight posterior distributions crossing (or being near) zero. In contrast, these replication cohorts offer increased evidence in support of an associational effect on SA (all posterior distributions above or nearly above zero) and offer increased support for the null hypothesis for the additional four global neuroimaging measures (CT, TBV, wFA, WM_hyper).

Supplementary Table 1

Fuzzy RD Placebo Outcome results

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

[2] 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.

Supplementary Table 2

Fuzzy RD Global Neuroimaging Results

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

Supplementary Table 3

Fuzzy RD Uncorrected Global Neuroimaging Results

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

Supplementary Table 4

One month window Bayesian Analysis

Y	X	n	Estimate	Std Est.	95 Credible Interval	BF	*normal*(0, 0.5)	*normal*(0, 1.5)
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

***Sup. Table 4 Caption**:* 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".

Supplementary Table 5

One Month window Test of Covariates

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

***Sup. Table 5 Caption**:* 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).

Supplementary Table 6

Five Month window Bayesian Analysis

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

***Sup. Table 6 Caption**:* 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	Bayes Factor	ci_low	ci_high
sex	ROSLA	1972	-0.04	BF_{01}=12.30	-0.08	0.01
visit_day_correct	ROSLA	1972	107.29	BF_{01}=1.78	25.87	191.74
visit_day_correct^2	ROSLA	1972	384967.33	BF_{01}=3.35	57251.82	720507.66
headmotion	ROSLA	1305	0.01	BF_{01}=10.67	0.00	0.03
imaging_center_11026	ROSLA	1950	0.01	BF_{01}=34.58	-0.02	0.04
imaging_center_11027	ROSLA	1950	0.02	BF_{01}=28.85	-0.02	0.06
dMRI_25922_1	ROSLA	1203	-0.01	BF_{01}=33.09	-0.06	0.04
dMRI_25921_1	ROSLA	1246	0.03	BF_{01}=19.13	-0.02	0.08
dMRI_25928_1	ROSLA	1208	0.07	BF_{01}=1.10	0.02	0.11
imaging_center_11028	ROSLA	1950	0.02	BF_{01}=15.85	-0.01	0.04
imaging_center_11025	ROSLA	1950	-0.04	BF_{01}=6.20	-0.09	0.00

***Sup. Table 7 Caption**:* 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).

Supplementary Table 7

Five Month window Test of Covariates

Code and Data Availability

All code is publicly available (*https://github.com/njudd/eduBrain*), for a very similar more streamlined script see (*https://github.com/njudd/EduTelomere*). 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.

Additional information

Funding & Acknowledgments

Nicholas Judd is supported with eScience and Jacobs fellowships. Rogier Kievit is supported by a Hypatia fellowship from the RadboudUMC. This research has been conducted using the UK Biobank resource under application number 23668. We would like to thank Barbara Sakic, Barbara Franke, Andre Marquand, Jan Buitelaar, Nina Roth Mota, Janita Bralten, and Ward de Witte for help in navigating and accessing the UK Biobank. We would like to thank Prof. Michal Kolesár for his help in covariate correction, along with highlighting the great code base from Dr. Margot van de Weijer (*https://github.com/* margotvandeweijer/EA_causality).

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.

References

1. (1950) **Geneva Convention IV. (1950).**

2. (1989) **United Nations Convention on the Rights of the Child. (1989).**

3. Lövdén M., Fratiglioni L., Glymour M. M., Lindenberger 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**:c654

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**:101939

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. Seyedsalehi A., et al. (2023) **Educational attainment, structural brain reserve and Alzheimer's disease: a Mendelian randomization analysis** *Brain* **146**:2059–2074

26. Hart S. A., Little C., van Bergen E (2021) **Nurture might be nature: cautionary tales and proposed solutions** *NPJ Sci Learn* **6**

27. 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

28. Farah M. J (2017) **The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences** *Neuron* **96**:56–71

29. 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

30. 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

31. 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

32. Lee D. S., Lemieux T (2010) **Regression Discontinuity Designs in Economics** *J. Econ. Lit* **48**:281–355

33. Cattaneo M. D., Titiunik R. (2022) **Regression discontinuity designs** *Annu. Rev. Econom*

34. Angrist J. D., Krueger A. B (1991) **Does Compulsory School Attendance Affect Schooling and Earnings?** *Q. J. Econ* **106**:979–1014

35. Armstrong T. B., Kolesár M (2018) **Optimal inference in a class of regression models** *Econometrica* **86**:655–683

36. Armstrong T. B., Kolesár M (2020) **Simple and honest confidence intervals in nonparametric regression** *Quant. Econom* **11**:1–39

37. Cattaneo M. D., Idrobo N., Titiunik R. (2019) **A practical introduction to regression discontinuity designs: Foundations** *arXiv* https://doi.org/10.48550/arXiv.1911.09511

38. (1972) **The Raising of the School Leaving Age Order**

39. 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

40. Gorman E (2023) **Does Schooling Have Lasting Effects on Cognitive Function? Evidence From Compulsory Schooling Laws** *Demography* **60**:1139–1161

41. Clark D., Royer H (2010) **The Effect of Education on Adult Health and Mortality: Evidence from Britain** *NBER Working Paper Series* https://doi.org/10.3386/w16013

42. Clark D., Royer H (2013) **The Effect of Education on Adult Mortality and Health: Evidence from Britain** *Am. Econ. Rev* **103**:2087–2120

43. Barcellos S. H., Carvalho L. S., Turley P (2023) **Distributional effects of education on health** *J. Hum. Resour* **58**:1273–1306

44. 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

45. 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

46. Fry A., et al. (2017) **Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population** *Am. J. Epidemiol* **186**:1026–1034

47. van Alten S., Domingue B. W., Faul J., Galama T., Marees A. T (2024) **Reweighting UK Biobank corrects for pervasive selection bias due to volunteering** *Int. J. Epidemiol* **53**:dyae054

48. Lyall D. M., et al. (2022) **Quantifying bias in psychological and physical health in the UK Biobank imaging sub-sample** *Brain Commun* **4**:fcac119

49. Geruso M., Royer H (2018) **The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK** *NBER Working Paper Series* https://doi.org/10.3386/w24332

50. Cattaneo M. D., Jansson M., Ma X (2018) **Manipulation Testing Based on Density Discontinuity** *Stata J* **18**:234–261

51. 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

52. de Groot M., et al. (2013) **Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration** *Neuroimage* **76**:400–411

53. Cattaneo M. D., Idrobo N., Titiunik R (2023) **A Practical Introduction to Regression Discontinuity Designs: Extensions** *arXiv* https://doi.org/10.48550/arXiv.2301.08958

54. Liu S., Abdellaoui A., Verweij K. J. H., van Wingen G. A (2023) **Replicable brain-phenotype associations require large-scale neuroimaging data** *Nat Hum Behav* https://doi.org/10.1038/s41562-023-01642-5

55. Jeffreys H. (1961) **Theory of probability (3rd edt.)** Oxford University Press

56. Draganski B., et al. (2006) **Temporal and spatial dynamics of brain structure changes during extensive learning** *J. Neurosci* **26**:6314–6317

57. Draganski B., et al. (2004) **Neuroplasticity: changes in grey matter induced by training** *Nature* **427**:311–312

58. Bråthen A. C. S., et al. (2022) **Cognitive and hippocampal changes weeks and years after memory training** *Sci. Rep* **12**:7877

59. 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

60. Garzón B., et al. (2023) **Cortical changes during the learning of sequences of simultaneous finger presses** *Imaging Neuroscience* **1**:1–26

61. Hille M., Kühn S., Kempermann G., Bonhoeffer T., Lindenberger U (2024) **From animal models to human individuality: Integrative approaches to the study of brain plasticity** *Neuron* **112**:3522–3541

62. Mediavilla T., et al. (2022) **Learning-related contraction of gray matter in rodent sensorimotor cortex is associated with adaptive myelination** *eLife* **11**

63. Schmidt S., et al. (2021) **Experience-dependent structural plasticity in the adult brain: How the learning brain grows** *Neuroimage* **225**

64. Goriounova N. A., et al. (2018) **Large and fast human pyramidal neurons associate with intelligence** *eLife* **7**

65. 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**

66. 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

67. López-Otín C., Blasco M. A., Partridge L., Serrano M., Kroemer G (2023) **Hallmarks of aging: An expanding universe** *Cell* **186**:243–278

68. Nyberg L., et al. (2021) **Educational attainment does not influence brain aging** *Proc. Natl. Acad. Sci. U. S. A* **118**

69. Burt S. A., Johnson W (2023) **Joint Consideration of Means and Variances Might Change the Understanding of Etiology** *Perspect. Psychol. Sci* **18**:416–427

70. von Stumm S., d'Apice K. (2021) **From Genome-Wide to Environment-Wide: Capturing the Environome** *Perspect. Psychol. Sci* **1745691620979803**

71. 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

72. 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

73. 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

74. 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

75. Munafò M. R., Tilling K., Taylor A. E., Evans D. M., Davey Smith G (2018) **Collider scope: when selection bias can substantially influence observed associations** *Int. J. Epidemiol* **47**:226–235

76. Choi J.-Y., Lee M.-J (2018) **Relaxing conditions for local average treatment effect in fuzzy regression discontinuity** *Econ. Lett* **173**:47–50

77. 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**:2624

78. Fischl B., et al. (2002) **Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain** *Neuron* **33**:341–355

79. McCarthy P (2020) **Funpack** *Zenodo* https://doi.org/10.5281/zenodo.3761702

80. Calonico S., Cattaneo M. D., Titiunik R (2014) **Robust nonparametric confidence intervals for regression-discontinuity designs** *Econometrica* **82**:2295–2326

81. van Buuren S., Groothuis-Oudshoorn K. (2011) **mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software** *Articles* **45**:1–67

82. Goodrich B., Gabry J., Ali I., Brilleman S. (2024) **rstanarm: Bayesian applied regression modeling via Stan** https://mc-stan.org/rstanarm/

83. , version 2.32. Preprint at (2023) **Stan Development Team. Stan Modeling Language Users Guide and Reference Manual** https://mc-stan.org

84. Depaoli S., van de Schoot R (2017) **Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist** *Psychol. Methods* **22**:240–261

85. Makowski D., Ben-Shachar M. S (2019) **bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework** *Journal of Open Source*

86. Judd N., et al. (2020) **Cognitive and brain development is independently influenced by socioeconomic status and polygenic scores for educational attainment** *Proc. Natl. Acad. Sci. U. S. A* **117**:12411–12418

87. Ritchie S. J., et al. (2018) **Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants** *Cereb. Cortex* **28**:2959–2975

88. Avendano M., de Coulon A., Nafilyan V (2020) **Does longer compulsory schooling affect mental health? Evidence from a British reform** *J. Public Econ* **183**

89. Smith Stephen, Almagro Fidel Alfaro, Miller Karla (no date) **UK Biobank Brain Imaging Documentation** https://mc-stan.org

Author information

Nicholas Judd
Cognitive Neuroscience Department, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, Netherlands
ORCID iD: 0000-0002-0196-9871

For correspondence: nicholas.judd@radboudumc.nl

Rogier Kievit
Cognitive Neuroscience Department, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, Netherlands
ORCID iD: 0000-0003-0700-4568

Editors

Reviewing Editor
Jason Lerch
University of Oxford, Oxford, United Kingdom

Senior Editor
Timothy Behrens
University of Oxford, Oxford, United Kingdom

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 anlaysis 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 anlayses).

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 measured long after the intervention. Thus a null finding here is completely consistent educational attainement (EA) in fact having an impact on brain structure, where EA may reflect elements of training after second education (e.g. university, post-graduate qualifications, etc) and not just stopping education at 16 yrs yes/no.

https://doi.org/10.7554/eLife.101526.2.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:

- 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.
- 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.
- 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.
- 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.
- 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:

- This study investigates a well-posed but necessarily narrow question in a specific setting: 15-year old British students born around 1957 who also participate 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.
- 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. However, other studies have not found specific issues that would invalidate ROSLA as a natural experiment.

https://doi.org/10.7554/eLife.101526.2.sa1

Author response:

The following is the authors' response to the original reviews

> ***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 strengthed from additional consideration from a methodological perspective.*

We sincerely thank the reviewer for the useful input, in particular, their recommendation to clarify RD and for catching some minor errors in the methods (such as taking the log of the Bayes factors).

> ***Reviewer #1 (Recommendations for the authors):***
>
> *(1) The fuzzy local-linear regression discontinuity analysis would benefit from further description.*
>
> *(2) In the description of the model, the terms "smoothness" and "continuity" appear to be used interchangeably. This should be adjusted to conform to mathematical definitions.*

We have now added to our explanations of continuity regression discontinuity. In particular, we now explain “fuzzy”, and add emphasis on the two separate empirical approaches (continuity and local-randomization), along with fixing our use of “smoothness” and “continuity”.

results:

“Compliance with ROSLA was very high (near 100%; Sup. Figure 2). However, given the cultural and historical trends leading to an increase in school attendance before ROSLA, most adolescents were continuing with education past 15 years of age before the policy change (Sup Plot. 7b). Prior work has estimated 25 percent of children would have left school a year earlier if not for ROSLA 41. Using the UK Biobank, we estimate this proportion to be around

10%, as the sample is healthier and of higher SES than the general population (Sup. Figure 2; Sup. Table 2) 46–48."

methods:

"RD designs, like ours, can be 'fuzzy' indicating when assignment only increases the probability of receiving it, in turn, treatment assigned and treatment received do not correspond for some units 33,53. For instance, due to cultural and historical trends, there was an increase in school attendance before ROSLA; most adolescents were continuing with education past 15 years of age (Sup Plot. 7b). Prior work has estimated that 25 percent of children would have left school a year earlier if not for ROSLA 41. Using the UK Biobank, we estimate this proportion to be around 10%, as the sample is healthier and of higher SES than the general population (Sup. Figure 2; Sup. Table 2) 46–48."

> *(3) The optimization of the smoother based on MSE would benefit from more explanation and consideration. How was the flexibility of the model taken into account in testing? Were there any concerns about post-selection inference? A sensitivity analysis across bandwidths is also necessary. Based on the model fit in Figure 1, results from a linear model should also be compared.*

It is common in the RD literature to illustrate plots with higher-order polynomial fits while inference is based on linear (or at most quadratic) models (Cattaneo, Idrobo & Titiunik, 2019). We agree that this field-specific practice can be confusing to readers. Therefore, we have redone Figure 1 using local-linear fits better aligning with our analysis pipeline. Yet, it is still not a *one-to-one* alignment as point estimation and confidence are handled robustly while our plotting tools are simple linear fits. In addition, we updated Sup. Fig 3 and moved 3rd-order polynomial RD plots to Sup. Fig 4.

Empirical RD has many branching analytical decisions (bandwidth, polynomial order, kernel) which can have large effects on the outcome. Fortunately, RD methodology is starting to become more standardized (Catteneo & Titiunik, 2022, Ann. Econ Rev) as there have been indications of publication bias using these methods (Stommes, Aronow & Sävje, 2023, Research and Politics (This paper suggest it is not researcher degrees of freedom, rather inappropriate inferential methods)). While not necessarily ill-intended, researcher degrees of freedom and analytic flexibility are major contributors to publication bias. We (self) limited our analytic flexibility by using pre-registration (https://osf.io/rv38z).

One of the most consequential analytic decisions in RD is the bandwidth size as there is no established practice, they are context-specific and can be highly influential on the results. The choice of bandwidths can be framed as a 'bias vs. variance trade-off'. As bandwidths increase, variance decreases since more subjects are added yet bias (misspecification error/smoothing bias) also increases (as these subjects are further away and less similar). In our case, our assignment (running/forcing) variable is 'date of birth in months'; therefore our smallest comparison would be individuals born in August 1957 (unaffected/no treatment) vs September 1957 (affected/treated). This comparison has the least bias (subjects are the most similar to each other), yet it comes at the expense of *very few* subjects (high variance in our estimate).

MSE-derived bandwidths attempt to solve this issue by offering an automatic method to choose an analysis bandwidth in RD. Specifically, this aims to minimize the MSE of the local polynomial RD point estimator – effectively choosing a bandwidth by balancing the 'bias vs. variance trade-off' (explained in detail 4.4.2 Cattaneo et al., 2019 p 45 - 51 "A practical introduction to regression discontinuity designs: foundations"). Yet, you are very correct in highlighting potential overfitting issues as they are "by construction invalid for inference" (Calonico, Cattaneo & Farrell, 2020, p. 192). Quoting from Cattaneo and Titiunik's Annual Review of Economics from 2022:

"Ignoring the misspecification bias can lead to substantial overrejection of the null hypothesis of no treatment effect. For example, back-of-the-envelop calculations show that a nominal 95% confidence interval would have an empirical coverage of about 80%."

Fortunately, modern RD analysis packages (such as rdrohust or RDHonest) calculate robust confidence intervals - for more details see Armstrong and Kolesar (2020). For a summary on MSE-bandwidths see the section "Why is it hard to estimate RD effects?" in Stommes and colleagues 2023 (https://arxiv.org/abs/2109.14526). For more in-depth handling see the Catteneo, Idrobo, and Titiunik primer (https://arxiv.org/abs/1911.09511).

Lastly, with MSE-derived bandwidths sensitivity tests only make sense within a narrow window of the MSE-optimized bandwidth (5.5 Cattaneo et al., 2019 p 106 - 107). When a significant effect occurs, placebo cutoffs (artificially moving the cutoff) and donut-hole analysis are great sensitivity tests. Instead of testing our bandwidths, we decided to use an alternate RD framework (local randomization) in which we compare 1-month and 5-month windows. Across all analysis strategies, MRI modalities, and brain regions, we do not find any effects of the education policy change ROSLA on long-term neural outcomes.

> *(4) In the Bayesian analysis, the authors deviated from their preregistered analytic plan. This whole section is a bit confusing in its current form - for example, point masses are not wide but rather narrow. Bayes factors are usually estimated; it is unclear how or why a prior was specified. What exactly is being modeled using a prior? Also, throughout - If the log was taken, as the methods seem to indicate for the Bayes factor, this should be mentioned in figures and reported estimates.*

First, we would like to thank you for spotting that we incorrectly kept the log in the methods. We have fixed this and added the following sentence to the methods:

"Bayes factors are reported as BF_{10} in support of the alternative hypothesis, we report Bayes factors under 1 as the multiplicative inverse (BF_{01} = 1/BF)"

All Bayesian analyses need to have a prior. In practice, this becomes an issue when you're uncertain about 1) the location of the effect (directionality & center mass, defined by a location parameter), yet more importantly, the 2) confidence/certainty of the range-spread of possible effects (determined by a scale parameter). In normally distributed priors these two 'beliefs' are represented with a mean and a standard deviation (the latter impacts your confidence/certainty on the range of plausible parameter space).

Supplementary figure 6 illustrates several distributions (location = 0 for all) with varying scale parameters; when used as Bayesian priors this indicates differing levels of confidence in our certainty of the plausible parameter space. We illustrate our three reported, normally distributed priors centered at zero in blue with their differing scale parameters (sd = .5, 1 & 1.5).

All of these five prior distributions have the same location parameter (i.e., 0) yet varying differences in the scale parameter – our confidence in the certainty of the plausible parameter space. At first glance it might seem like a flat/uniform prior (not represented) is a good idea – yet, this would put equal weight on the possibility of every estimate thereby giving the same probability mass to implausible values as plausible ones. A uniform prior would, for instance, encode the hypothesis that education causing a 1% increase in brain volume is just as plausible as it causing either a doubling or halving in brain volume. In human research, we roughly know a range of reasonable effect sizes and it is rare to see massive effects.

A benefit of 'weakly-informative' priors is that they limit the range of plausible parameter values. The default prior in STAN (a popular Bayesian estimation program; https://mc-stan

.org) is a normally distributed prior with a mean of zero and an SD of 2.5 (seen in orange in the figure; our initial preregistered prior). This large standard deviation easily permits positive and negative estimates putting minimal emphasis on zero. Contrast this to BayesFactor package's (Morey R, Rouder J, 2023) default "wide" prior which is the Cauchy distribution (0, .7) illustrated in magenta (for more on the Cauchy see: https://distribution-explorer.github.io/continuous/cauchy.html).

These different defaults reflect differing Bayesian philosophical schools ('estimate parameters' vs 'quantify evidence' camps); if your goal is to accurately estimate a parameter it would be odd to have a strong null prior, yet (in our opinion) when estimating point-null BF's a wide default prior gives far too much evidence in support of the null. In point-null BF testing the Savage-Dickey density ratio is the ratio between the height of the prior at 0 and the height of the posterior at zero (see Figure under section "testing against point null 0"). This means BFs can be very prior sensitive (seen in SI tables 5 & 6). For this reason, we thought it made sense to do prior sensitivity testing, to ensure our conclusions in favor of the null were not caused solely by an overly wide prior (preregistered orange distribution) we decided to report the 3 narrower priors (blue ones).

Alternative Bayesian null hypotheses testing methods such as using Bayes Factors to test against a null region and 'region of practical equivalence testing' are less prior sensitive, yet both methods demand the researcher (e.g. 'us') to decide on a minimal effect size of practical interest. Once a minimal effect size of interest is determined any effect within this boundary is taken as evidence in support of the null hypothesis.

> *(5) It is unclear why a different method was employed for the August / September data analysis compared to the full-time series.*

We used a local-randomization RD framework, an entirely different empirical framework than continuity methods (resulting in a different estimate). For an overview see the primer by Cattaneo, Idrobo & Titiunik 2023 ("A Practical Introduction to Regression Discontinuity Designs: Extensions"; https://arxiv.org/abs/2301.08958).

A local randomization framework is optimal when the running variable is discrete (as in our case with DOB in months) (Cattaneo, Idrobo & Titiunik 2023). It makes stronger assumptions on exchangeability therefore a very narrow window around the cutoff needs to be used. See Figure 2.1 and 2.2 (in the Cattaneo, Idrobo & Titiunik 2023) for graphical illustrations of 1) a randomized experiment, 2) a continuity RD design, and 3) local-randomization RD. Using the full-time series in a local randomization analysis is not recommended as there is no control for differences between individuals as we move further away from the cutoff – making the estimated parameter highly endogenous.

We understand how it is confusing to have both a new framework and Bayesian methods (we could have chosen a fully frequentist approach) but using a different framework allows us to weigh up the aforementioned 'bias vs variance tradeoff' while Bayesian methods allow us to say something about the weight of evidence (for or against) our hypothesis.

> *(6) Figure 1 - why not use model fits from those employed for hypothesis testing?*

This is a great suggestion (ties into #3), we have now redone Figure 1.

> *(7) The section on "correlational effect" might also benefit from additional analyses and clarifications. Indeed, the data come from the same randomized experiment for which minimum education requirements were adjusted. Was the only difference that the number of years of education was studied as opposed to the cohort? If so, would the*

> *results of this analysis be similar in another subsample of the UK Biobank for which there was no change in policy?*

We have clarified the methods section for the correlational/associational effect. This was the same subset of individuals for the local randomization analysis; all we did was change the independent variable from an exogenous dummy-coded ROSLA term (where half of the sample had the natural experiment) to a continuous (endogenous) educational attainment IV.

In principle, the results from the associational analysis should be exactly the same if we use other UK Biobank cohorts. To see if the association of education attainment with the global neuroimaging cohorts was similar across sub-cohorts of new individuals, we conducted post hoc Bayesian analysis on eight more subcohort of 10-month intervals, spaced 2 years apart from each other (Sup. Figure 7; each indicated by a different color). Four of these sub-cohorts predate ROSLA, while the other four are after ROSLA. Educational attainment is slowly increasing across the cohorts of individuals born from 1949 until 1965; intriguingly the effect of ROSLA is visually evident in the distributions of educational attainment (Sup. Figure 7). Also, as seen in the cohorts predating ROSLA more and more individuals were (already) choosing to stay in education past 15 years of age (see cohort 1949 vs 1955 in Sup. Figure 7).

Sup. Figure 8 illustrates boxplots of the educational attainment posterior of the eight sub-cohorts in addition to our original analysis (s1957) using a normal distributed prior with a mean of 0 and a sd of 1. Total surface area shows a remarkably replicable association with education attainment. Yet, it is evident the "extremely strong" association we found for CSF was a statistical fluke – as the posterior of other cohorts (bar our initial test) crosses zero. The conclusions for the other global neuroimaging covariates where we concluded 'no associational effect' seems to hold across cohorts.

We have now added methods, deviation from preregistration, and the following excerpt to the results:

"A post hoc replication of this associational analysis in eight additional 10-month cohorts spaced two years apart (Sup. Figure 7) indicates our preregistered report on the associational effect of educational attainment on CSF to be most likely a false-positive (Sup. Figure 8). Yet, the positive association between surface area and educational attainment is robust across the additional eight replication cohorts."

> ***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., Boultwood, 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*

We thank the reviewer for the positive comments and constructive feedback, in particular, their emphasis on volunteer bias in UKB (similar points were mentioned by Reviewer 3). We have now addressed these limitations with the following passage in the discussion:

"The UK Biobank is known to have 'healthy volunteer bias', as respondents tend to be healthier, more educated, and are more likely to own assets [71,72]. Various types of selection bias can occur in non-representative samples, impacting either internal (type 1) or external (type 2) validity. One benefit of a natural experimental design is that it protects against threats to internal validity from selection bias [43], design-based internal validity threats still exist, such as if volunteer bias differentially impacts individuals based on the cutoff for assignment. A more pressing limitation – in particular, for an education policy change – is our power to detect effects using a sample of higher-educated individuals. This is evident in our first stage analysis examining the percentage of 15-year-olds impacted by ROSLA, which we estimate to be 10% in neuro-UKB (Sup. Figure 2 & Sup. Table 2), yet has been reported to be 25% in the UK general population [41]. Our results should be interpreted for this subpopulation (UK, 1973, from 15 to 16 years of age, compliers) as we estimate a 'local' average treatment effect [73]. Natural experimental designs such as ours offer the potential for high internal validity at the expense of external validity."

We also highlighted it both in the results and methods.

We appreciate that one year of education may seem modest compared to the entire educational trajectory, but as an intervention, we disagree that one year of education is 'a very modest manipulation'. It is arguably one of the largest positive manipulations in childhood development we can administer. If we were to translate a year of education into the language of a (cognitive) intervention, it is clear that the manipulation, at least in terms of hours, days, and weeks, is substantial. Prior work on structural plasticity (e.g., motor, spatial & cognitive training) has involved substantially more limited manipulations in time, intensity, and extent. There is even (limited) evidence of localized persistent long-term structural changes (Wollett & Maguire, 2011, Cur. Bio.).

We have now also highlighted the limited generalizability of our findings since we estimate a 'local' average treatment effect. It is possible higher education (college, university, vocational

schools, etc.) could impact brain structure, yet we see no theoretical reason why it would while secondary wouldn't. Moreover, higher education education is even trickier to research empirically due to heightened self and administrative selection pressures. While we cannot discount this possibility, the impacts of endogenous factors such as genetics and socioeconomic status are most likely heightened. That being said, higher education offers exciting possibilities to compare more domain-specific processes (e.g., by comparing a philosophy student to a mathematics student). Causality could be tested in European systems with point entry into *field-specific* programs – allowing comparison of students who just missed entry criteria into one topic and settled for another.

Regarding the amount of time following the manipulation, as we highlight in our discussion this is both a weakness and a strength. Viewed from a developmental neuroplasticity lens it would have been nice to have imaging immediately following the manipulation. Yet, from an aging perspective, our design has *increased power* to detect an effect.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) The authors assert there is no strong causal evidence for EA on brain structure. This overlooks work from Mendielian Randomisation, e.g. this careful work: https://pubmed.ncbi.nlm.nih.gov/36310536/ ... evidence from (good quality) MR studies should be considered.*

We thank the reviewer for highlighting this well-done mendelian randomization study. We have now added this citation and removed previous claims on the "lack of causal evidence existing". We refrain from discussing Mendelian randomization, as it it would need to be accompanied by a nuanced discussion on the strong limitations regarding EduYears-PGS in Mendelian randomization designs.

> *(2) Tukey/Boxplot is a good name for your identification of outliers but your treatment of outliers has a well-recognized name that is missing: Windsorisation. Please add this term to your description to help the reader more quickly understand what was done.*

Thanks, we have now added the term winsorized.

> *(3) Nowhere is it plainly stated that "fuzzy" means that you allow for imperfect compliance with the exposure, i.e. some children born before the cut-off stayed in school until 16, and some born after the cut-off left school before 16. For those unfamiliar with RD it would be very helpful to explain this at or near the first reference of the term "fuzzy".*

We have now clarified the term 'fuzzy' to the results and methods:

methods:

"RD designs, like ours, can be 'fuzzy' indicating when assignment only increases the probability of receiving it, in turn, treatment assigned and treatment received do not correspond for some units 33,53. For instance, due to cultural and historical trends, there was an increase in school attendance before ROSLA; most adolescents were continuing with education past 15 years of age (Sup Plot. 7b). Prior work has estimated that 25 percent of children would have left school a year earlier if not for ROSLA 41. Using the UK Biobank, we estimate this proportion to be around 10%, as the sample is healthier and of higher SES than the general population (Sup. Figure 2; Sup. Table 2) 46–48."

> *(4) Supplementary Figure 2 never states what the percentage actually measures. What exactly does each dot represent? Is it based on UK Biobank subjects with a given birth*

month? If so clarify.

Fixed!

Reviewer #3 (Public review):

Summary:

This study investigates evidence for a hypothesized, 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 (Lyall 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.*

We thank the reviewer for their numerous positive comments and have now attempted to address the first two limitations (generalizability and UKB bias) with the following passage in the discussion:

“The UK Biobank is known to have ‘healthy volunteer bias’, as respondents tend to be healthier, more educated, and are more likely to own assets [71,72]. Various types of selection bias can occur in non-representative samples, impacting either internal (type 1) or external (type 2) validity. One benefit of a natural experimental design is that it protects against threats to internal validity from selection bias [43], design-based internal validity threats still exist, such as if volunteer bias differentially impacts individuals based on the cutoff for assignment. A more pressing limitation – in particular, for an education policy change – is our power to detect effects using a sample of higher-educated individuals. This is evident in our first stage analysis examining the percentage of 15-year-olds impacted by ROSLA, which we estimate to be 10% in neuro-UKB (Sup. Figure 2 & Sup. Table 2), yet has been reported to be 25% in the UK general population [41]. Our results should be interpreted for this subpopulation (UK, 1973, from 15 to 16 years of age, compliers) as we estimate a ‘local’ average treatment effect [73]. Natural experimental designs such as ours offer the potential for high internal validity at the expense of external validity.”

We further highlight this in the results section:

“Compliance with ROSLA was very high (near 100%; Sup. Figure 2). However, given the cultural and historical trends leading to an increase in school attendance before ROSLA, most adolescents were continuing with education past 15 years of age before the policy change (Sup Plot. 7b). Prior work has estimated 25 percent of children would have left school a year earlier if not for ROSLA 41. Using the UK Biobank, we estimate this proportion to be around 10%, as the sample is healthier and of higher SES than the general population (Sup. Figure 2; Sup. Table 2) 46–48.”

Healthy volunteer bias can create two types of selection bias; crucially participation *itself* can serve as a collider threatening internal validity (outlined in van Alten et al., 2024; https://academic.oup.com/ije/article/53/3/dyae054/7666749). Natural experimental designs are partially sheltered from this *major* limitation, as ‘volunteer bias’ would have to differentially

impact individuals on one side of the cutoff and not the other – thereby breaking a primary design assumption of regression discontinuity. Substantial prior work (including this article) has not found any threats to the validity of the 1973 ROSLA (Clark & Royer 2010, 2013; Barcellos et al., 2018, 2023; Davies et al., 2018, 2023). While the Davies 2028 article did IP-weight with the UK Biobank sample, Barcellos and colleagues 2023 (and 2018) do not, highlighting the following "Although the sample is not nationally representative, our estimates have internal validity because there is no differential selection on the two sides of the September 1, 1957 cutoff – see Appendix A.".

The second (more acknowledged & arguably less problematic) type of selection bias results in threats to external validity (aka generalizability). As highlighted in your first point; this is a large limitation with *every* natural experimental design, yet in our case, this is further amplified by the UK Biobank's healthy volunteer bias. We have now attempted to highlight this limitation in the discussion passage above.

Point 3 – the inability to fully confirm design validity – is again, another inherent limitation of a natural experimental approach. That being said, extensive prior work has tested different predetermined covariates in the 1973 ROSLA (cited within), and to our knowledge, no issues have been found. The 1973 ROSLA seems to be one of the better natural experiments around (there was also a concerted effort to have an 'effective' additional year; see Clark & Royer 2010). For these reasons, we stuck with only testing the variables we wanted to use to increase precision (also offering new neuroimaging covariates that didn't exist in the literature base). One additional benefit of ROSLA was that the cutoff was decided years later on a variable that happened (date of birth) in the past – making it particularly hard for adolescents to alter their assignments.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *(1) FMRIB's preprocessing pipeline is mentioned. Does this include deconfounding of brain measures? Particularly, were measures deconfounded for age before the main analysis?*

This is such a crucial point that we triple-checked, brain imaging phenotypes were not corrected for age (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf) – large effects of age can be seen in the global metrics; older individuals have less surface area, thinner cortices, less brain volume (corrected for head size), more CSF volume (corrected for head size), more white matter hyperintensities, and worse FA values. Figure 1 shows these large age effects, which are controlled for in our continuity-based RD analysis.

One's date of birth (DOB) of course does not match perfectly to their age, this is why we included the covariate 'visit date'; this interplay can now be seen in our updated SI Figure 1 (recommended in #3) which shows the distributions of visit date, DOB, and age of scan.

In a valid RD design covariates *should* not be necessary (as they should be balanced on either side of the cutoff), yet the inclusion of covariates does increase precision to detect effects. We tested this assumption, finding the effect of 'visit date' and its quadratic term to be not related to ROSLA (Sup. Table 1). This adds further evidence (specific to the UK Biobank sample) to the existing body of work showing the 1973 ROSLA policy change to not violate any design assumptions. Threats to internal validity would more than likely increase endogeneity and result in 'false causal positive causal effects' (which is not what we find).

> *(2) Despite the large overall sample size, I am wondering whether the effective number of samples is sufficient to detect a potentially subtle effect that is further attenuated by the long time interval before scanning. As stated, for the optimised bandwidth window (DoB 20 to 35 months around cut-off), N is about 5000. Does this mean that effectively about 250 (10%) out of about 2500 participants born after the cut-off were leaving school at 16*

> *rather than 15 because of ROSLA? For the local randomisation analysis, this becomes about N=10 (10% out of 100). Could a power analysis show that these cohort sizes are large enough to detect a reasonably large effect?*

This is a very valid point, one which we were grappling with while the paper was out for review. We now draw attention to this in the results and highlight this as a limitation in the discussion. While UKB's non-representativeness limits our power (10% affected rather than 25% in the general population), it is still a very large sample. Our sample size is more in line with standard neuroimaging studies than with large cohort studies.

The novelty of our study is its causal design, while we could very precisely measure an effect of some phenotype (variable X) in 40,000 individuals. This effect is probably not what we think we are measuring. Without IP-weighting it could even have a different sign. But more importantly, it is not variable X – it is the thousands of things (unmeasured confounders) that lead an individual to have more or less of variable X. The larger the sample the easier it is for small unmeasured confounders to reach significance (Big data paradox) – this in no way invalidates large samples, it is just our thinking and how we handle large samples will hopefully change to a more casual lens.

> *(3) Supplementary Figure 1: A similar raincloud plot of date of birth would be instructive to visualise the distribution of subjects born before and after the 1957 cut-off.*

Great idea! We have done this in Sup Fig. 1 for both visit date and DOB.

> *(4) p.9: Not sure about "extreme evidence", very strong would probably be sufficient.*

As preregistered, we interpreted Bayes Factors using Jeffrey's criteria. 'Extreme evidence' is only used once and it is about finding an associational effect of educational attainment on CSF (BF10 > 100). Upon Reviewer 1's recommendation 7, we conducted eight replication samples (Sup. Figure 7 & 8) and have now added the following passage to the results:

"A post hoc replication of this associational analysis in eight additional 10-month cohorts spaced two years apart (Sup. Figure 7) indicates our preregistered report on the associational effect of educational attainment on CSF to be most likely a false-positive (Sup. Figure 8). Yet, the positive association between surface area and educational attainment is robust across the additional eight replication cohorts."

> *(5) The code would benefit from a bit of clean-up and additional documentation. In its current state, it is not easy to use, e.g. in a replication study.*

We have now further added documentation to our code; including a readme describing what each script does. The analysis pipeline used is not ideal for replications as the package used for continuity-based RD (RDHonest) initially could not handle covariates – therefore we manually corrected our variables after a discussion with Prof Kolesár (https://github.com/kolesarm/RDHonest/issues/7).

Prof Kolesár added this functionality recently and future work should use the latest version of the package as it can correct for covariates. We have a new preprint examining the effect of 1972 ROLSA on telomere length in the UK Biobank using the latest package version of RDHonest (https://www.biorxiv.org/content/10.1101/2025.01.17.633604v1). To ensure maximum availability of such innovations, we will ensure the most up-to-date version of this script becomes available on this GitHub link (https://github.com/njudd/EduTelomere).

https://doi.org/10.7554/eLife.101526.2.sa0