



Are bigger brains smarter? Evidence from a large-scale pre-registered study

Journal:	<i>Psychological Science</i>
Manuscript ID	PSCI-17-1601.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	24-Jul-2018
Complete List of Authors:	Nave, Gideon; University of Pennsylvania Wharton School, Marketing Jung, Wi Hoon ; Korea University, Psychology Linnér, Richard; Vrije Universiteit Amsterdam Faculteit Economische wetenschappen en Bedrijfskunde, Economics Kable, Joseph; University of Pennsylvania, Psychology Koellinger, Philipp; Vrije Universiteit Amsterdam Faculteit Economische wetenschappen en Bedrijfskunde, Economics
Keywords:	Cognitive Ability, Individual Differences, Brain volume, Pre-registered analysis, UK Biobank

Are bigger brains smarter? Evidence from a large-scale pre-registered study

Gideon Nave¹, Wi Hoon Jung^{2,3}, Richard Karlsson Linnér⁴, Joseph Kable^{1,3}, Philipp D. Koellinger⁴

1 Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104, USA

2 Department of Psychology, Korea University, Seoul 136-701, South Korea

3 Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104, USA

4 Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

Corresponding author: Philipp D. Koellinger (p.d.koellinger@vu.nl)

Abstract

A positive relationship between brain volume and intelligence has been suspected since the 19th century and empirical studies seem to support this hypothesis. However, this claim is not uncontroversial because of concerns about publication bias and the lack of systematic control for critical confounding factors (e.g., height, population structure). We performed a pre-registered study of the relationship between brain volume and cognitive performance in a new adult population sample from the UK that is about 70% larger than all previous investigations on this subject combined ($N=13,608$). Our analyses systematically controlled for sex, age, height, socioeconomic status and population structure, and is free of publication bias. We find a robust association between total brain volume and fluid intelligence ($r=0.19$), which is consistent with previous findings in the literature after controlling for measurement quality of intelligence in our data. We also find a positive relationship with educational attainment ($r=0.12$). These relationships are mainly driven by grey matter (rather than white matter or fluid volume) and effect sizes are similar for both sexes and across age groups.

Keywords

Intelligence, educational attainment, brain volume, pre-registered analysis, UK Biobank

Article text

From logical reasoning to grasping new concepts, humans differ in cognitive capacities. A substantial part of this variance is captured by psychometric measures such as fluid intelligence tests or the general intelligence factor (g) which aggregates test results across various domains of cognitive performance. These measures are reliable, stable across the lifespan (Deary et al. 2000), and are associated with important life outcomes including educational attainment (Deary et al. 2007), job performance, and health (Batty et al. 2009).

Much research has been devoted to understanding how individual differences in cognitive performance arise, and whether they can be accounted for by environmental, developmental, genetic, and neuroanatomical factors. A classic hypothesis proposes a positive association between intelligence and total brain volume (TBV ; e.g., Francis Galton 1889). For decades, the only way to test this hypothesis were empirical studies using proxies of TBV such as head circumference. However, this work was controversial due to methodological issues (Stott 1983) and concerns about racial and cultural bias.

The introduction of Magnetic Resonance Imaging (MRI) in the late 1980s led to a burst of studies that directly examined the relationship between TBV and intelligence. The first published study reported a correlation of $r=.51$ in a sample of 40 college students (Willerman et al. 1991/4). However, the reported association has declined as sample sizes grew: the first meta-analysis of the literature ($k=14$, $N=858$) estimated an average correlation of $r=.37$ (G. Gignac, Vernon, and Wickett 2003). A later, more comprehensive meta-analysis ($k=37$, $N=1,530$) estimated a smaller correlation of $r=.29$ (McDaniel 2005/7). The largest meta-analysis to date

that included unpublished data reported an even smaller effect of $r=.24$ ($k=88$, $N=8,036$, (Pietschnig et al. 2015)).

Scholars have been debating the reliability, size, and meaning of a relationship between *TBV* and cognitive ability for many years (e.g., Stott 1983). Finding consensus is impeded by three main limitations.

First, only few studies systematically controlled for confounding factors such as height, age and socioeconomic status.

A second concern is population stratification, i.e. systematic biological differences across groups that might correlate with environmental and cultural factors.¹ If not properly controlled for, population stratification can induce a spurious relationship between biomarkers and phenotypes (Cardon and Palmer 2003). For example, individuals of north-west European descent may be slightly taller, have slightly larger brains, and perform slightly better in intelligence tests. But this effect could be primarily driven by more favorable environments (e.g. better schools, better healthcare) that could confound the relationship between *TBV* and intelligence. Genetic association studies have shown that self-reported ethnicity is often not sufficient to correct for such confounds. However, controlling for the first few principal components from the genetic data of the study participants has proven to be an effective strategy that is now standard in genetic association studies (Price et al. 2006; Rietveld, Conley, et al. 2014).

¹ Population stratification is a well-known concern in genetic association studies. For example, a spurious relationship between the LCT gene that codes for the enzyme lactase and EA is found if genetic association studies do not properly control for population stratification (Rietveld, Conley, et al. 2014). Lactose intolerance is unrelated to cognitive ability and is much more frequent in south-eastern parts of Europe than in north-western parts.

1
2
3
4
5 A third issue is a bias towards publication of positive, statistically significant results and effect
6 sizes that overestimate the true values. The most recent meta-analysis on intelligence and *TBV* by
7 Pietschnig et al. (2015) found evidence for publication bias and showed that the correlation in
8 *published* reports was $r=.30$ ($k=53$; $N=3,956$), but only $r=.17$ in a larger set of *unpublished*
9 studies ($k=67$; $N=2,822$). In contrast, (G. E. Gignac and Bates 2017) did not find evidence for
10 publication bias. However, their analysis was restricted to published studies of healthy
11 participants only. While several analytical techniques have been proposed to detect such bias,
12 their capacity to estimate the true effect size is controversial and their power to reject the null
13 hypothesis of no publication bias is low in small samples (Ioannidis et al. 2014). A clean
14 approach to avoid publication bias is to conduct a well-powered study following a pre-registered
15 analysis plan (Gonzales and Cunningham 2015).
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 We address these three shortcomings of the current literature here. Specifically, we conducted a
32 pre-registered analysis of the relationship between measures of cognitive performance and *TBV*
33 using data from the UK Biobank (hereandafter UKB; Miller et al. 2016; Sudlow et al. 2015). The
34 UKB is a data collection of unprecedented richness and scale that was not part of any previous
35 study on the relationship between *TBV* and cognitive performance. Our final sample contains
36 $N=13,608$ genotyped individuals with anatomical MRI brain scans, coming from an adult
37 population (>40 years old) of European decent, all of whom completed at least one test of
38 cognitive performance. This sample is $\approx 70\%$ larger than all previous studies associating in-vivo
39 *TBV* and intelligence combined (Pietschnig et al. 2015), it permits novel ways to control for
40 confounds, and allows comparing effect sizes across various demographic groups.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Our investigation provides the opportunity for two additional contributions. First, we investigate the differential contributions of grey matter (neuronal cell bodies, dendrites, unmyelinated axons, glial cells, synapses, and capillaries), white matter (myelinated axons, or tracts), and cerebrospinal fluid to the association between *TBV* and intelligence. Both grey and white matter volumes are genetically correlated with general intelligence (Sniers et al. 2017) and are thought to contribute to the association based on small sample studies (e.g., Haier et al. 2004); understanding their differential contributions is essential for further theoretical development of accounts for the relationship between *TBV* and intelligence.

Second, we examine the association between *TBV* and educational attainment (*EA*), an important real-life outcome that crucially impacts individuals' income, health, and longevity (Lager and Torssander 2012). To date, this association has only been investigated by a few small sample studies of elderly or clinical populations, e.g., (Coffey et al. 1999).

Methods

The UK Biobank data

The UKB (Miller et al. 2016; Sudlow et al. 2015) recruited 502,617 people aged between 40-69 years in 2006-2010 from the general population across the entire UK. Almost all participants (488,363) have been genotyped (Bycroft et al. 2017), and extensive batteries of lifestyle measures have already been collected. The project aims to acquire high-quality MRI scan data from 100,000 participants in the next few years (Miller et al. 2016), following a standardized protocol at three dedicated, identical scanning centers operating 7 days per week, each scanning 18 subjects per day (Petersen et al. 2013). As of April 2018, 15,040 participants have already

1
2
3 been scanned and their T1 structural brain images have been processed by the UKB team
4
5 (Stephen Smith 2014) and converted from DICOM to NIFTI format. Health outcomes are
6
7 tracked over time for all participants by linking the Biobank to official hospital records. The
8
9 principle goal of the project is to use large-scale longitudinal data in order to better understand
10
11 disease etiology and to develop predictive methods for early onset disease detection. An
12
13 important byproduct of the Biobank project is the generation of an unprecedentedly large and
14
15 rich dataset to study behavioral phenotypes and their relation to the collected biological markers
16
17 (e.g., genotypes, brain scans) and health outcomes (e.g. cognitive performance, subjective well-
18
19 being, BMI, diseases).
20
21
22
23
24
25
26

27 Measures

30 *Fluid intelligence*

31
32
33 The UKB contains a short measure of verbal-numerical reasoning (referred to as “fluid
34
35 intelligence test”) that consists of 13 multiple-choice questions (see **Supplemental Material**)
36
37 measuring the capacity to solve problems that require logic and reasoning ability, independent of
38
39 acquired knowledge. Participants had two minutes to complete as many questions as possible
40
41 from the test. The fluid intelligence test score is the simple unweighted sum of the number of
42
43 correct answers given to these 13 questions. Participants who did not answer all of the questions
44
45 within the allotted 2-minute limit are scored as zero for each of the unattempted questions.
46
47
48
49
50
51

52 The fluid intelligence test was administered on three occasions: (1) the initial assessment visit,
53
54 (2) the first repeat assessment visit, and (3) the imaging visit (see below). The test was also
55
56
57
58
59
60

administered in an online follow-up, which contained one additional question (thus, the maximum score was 14). The pairwise correlation between measurement instances in the sample that included brain scans and genotypes was between .60 and .69 (N between 989 and 7,584, see **Table S1**), consistent with earlier reports (Lyall et al. 2016). Participants did not receive feedback about their performance and they were not informed about the correct answers to the test questions at any point in time. We had access to $N=14,021$ with brain scans and at least one measurement instance of fluid intelligence. To maximize sample size and to reduce noise in the measure, we aggregated the scores of all measurement instances. To do so, we standardized each score separately to have a mean of 0 and a standard deviation of 1. We constructed the variable *fluid intelligence* of each participant by taking the average of these standardized scores (in cases multiple observations were available for an individual), and standardized the resulting measure again. To control for differences among individuals who participated in different test instances (e.g., participants who have taken all four tests vs. only one of them), we generated indicator variables for each one of the tests (i.e., a variable that equals one if the participant took a specific instance of the test and zero otherwise, and likewise for the other test instances) and included them as control variables in the regression analyses.

Other cognitive measures

Apart from the fluid intelligence measure, we performed robustness checks and additional exploratory analyses using three additional cognitive tests that are currently available in a large subsample of the UKB (numeric memory, reaction times, and visual memory). The psychometric properties of these tests are described in detail in Lyall et al. (2016).

Numeric memory was measured by a task that first showed participants a two-digit number and

asked to recall that number after a short pause. The number of digits then increased by one until either an error was made or the maximum number of twelve digits was reached, and the final number of digits shown was recorded. A higher number implies better cognitive performance. In the reaction time task, participants completed a timed test of symbol matching similar to the known card game ‘Snap’ and their mean response time across trials containing matching pairs was recorded. Higher scores imply slower responses, i.e. lower cognitive performance. Visual memory was measured by a task in which participants memorized the positions of either 3 or 6 card pairs and then had to match them from memory while making as few errors as possible. The test score denotes the number of errors made (i.e., higher scores imply lower cognitive performance).

General cognitive ability (‘g’)

It is well-known that low measurement quality can attenuate the estimated relationship between variables (Greene 2003) and G. E. Gignac and Bates (2017) find substantially higher correlations between brain size and cognitive ability in studies with “excellent” measures of IQ than in studies with “good” or “fair” measures (.39 [95% CI: .32, .46], .32 [.16, .46], and .21 [.14, .28], respectively). As a robustness check of our main results based on the crude fluid intelligence test described above, we repeated our analysis using four more comprehensive measures of general cognitive ability (henceforth referred to as ‘g’). Our measures of g used the fluid intelligence test as well as the three additional cognitive tests available in the UKB, described above.

Our primary measure of g employed all available measurement instances of these tests and standardized each instance separately. Then, we averaged across instances and standardized the

resulting measure again. Following standard practice in the literature, we extracted the first unrotated principal component from these various measures of cognitive performance to obtain a proxy for g (Lyall et al. 2016; Rietveld, Esko, et al. 2014; Benyamin et al. 2013), yielding $N=7,511$.

Consistent with earlier studies, we find that *fluid intelligence* has the highest loading on g -- .77 in Lyall et al. (2016) and .78 in our data, see **Table S2**. In our analyses, we chose to focus on *fluid intelligence* instead of g because (i) the numeric memory test is only available in a subset of our participants which reduces the sample size for g analyses by almost 50% compared to *fluid intelligence* and (ii) imputation of missing observations is not possible without potentially introducing substantial noise (Rubin 2004). We preferred *fluid intelligence* over the other two cognitive tests that were available in our entire sample (reaction time and visual memory) because these two have substantially lower loadings on g (-.37 and -.48, respectively) and lower retest reliability ($r_{reaction_time} \approx .55$, $r_{visual_memory} \approx .21$, see **Table S1**).

Our second measure of g was constructed by performing factor analysis (FA) of a single factor on the four tests, instead of principal component analysis. The analysis used minimum residuals estimation and oblimin rotation. The correlation between this measure of g and our primary measure derived from principal component analysis (PCA) was $r_{g-FA, g-PCA} = .94$.

Our third measure of g used a previously published protocol to construct g in the UKB described by Lyall et al. (2016). This protocol only made use of the data from the first touchscreen interview, it ignored data from the 3-pair version of the pair matching test, and used LN-

transformations of reaction time and $\text{LN} + 1$ of the visual memory tests. It then used PCA as a data reduction technique to extract g ($N=1,017$).

Our fourth measure of g was a general cognitive ability measurement constructed in a similar manner to our primary measure, but now excluding the fluid intelligence scores before performing PCA ($N=7,511$). This provides a measure of g that is independent from our main *fluid intelligence* measure.

All of these four measures of intelligence would be rated as “good” according to the guidelines established by G. E. Gignac and Bates (2017) (i.e. 2-8 tests, 2-3 dimensions, 20-39 min testing time), compared to a “poor” rating of our main measure of *fluid intelligence* (one test, one dimension, very short testing time). However, *fluid intelligence* allows us to study the relationship with *TBV* in a substantially larger sample ($N=13,608$ versus $N=1,017$ to $N=7,511$).

Educational Attainment (EA)

Following the standard established by the Social Science Genetic Association Consortium (SSGAC) (Rietveld et al. 2013), we measure *EA* as US years-of-schooling equivalents for the highest educational degree an individual obtained. We follow the ISCED 1997 classification (UNESCO), which leads to seven categories of EA that are internationally comparable. EA was measured via self-reports in the UKB on 3 occasions: (1) the initial assessment visit, (2) the first repeat assessment visit, and (3) the imaging visit. We use the highest educational degree reported on any of these occasions as our measure of EA.

Total brain volume

The UKB collected T1-weighted structural brain images using a 3-T Siemens Skyra with a 32-channel head coil (Siemens, Erlangen, Germany). The scanning parameters were as follows: repetition time (TR)=2000 ms; echo time (TE)=2.1 ms; flip angle=8°; matrix size=256 × 256 mm; voxel size=1 × 1 × 1 mm; and number of slices=208. Instead of using the preprocessed brain size variables provided by the UKB, we analyzed the T1-weighted images ourselves with the Computational Anatomy Toolbox (CAT12) implemented in SPM12.² The CAT12 software is a fully automated toolbox for measurements of gray matter (*GM*) and white matter (*WM*) volumes and cortical thickness at voxel and region-of-interest levels. Image preprocessing used the default settings of CAT12. Images were corrected for bias-field inhomogeneity, segmented into *GM*, *WM*, and cerebrospinal fluid (*CSF*), spatially normalized to the MNI space using linear and nonlinear transformations, and modulated to preserve the total amount of signal in the original image during spatial normalization. Total brain volume (*TBV*) was calculated by summing the raw volumes of *GM*, *WM*, and *CSF*.

For quality assurance, we conducted the following checks. First, we inspected all T1 images that were available to us as of April 2018 visually ($N=14,793$) and excluded 48 images due to artifacts, poor image quality, or gross brain pathology hampering image segmentation. Next, we processed the images using the CAT12 toolkit (Gaser and Dahnke 2016) and performed the sample homogeneity check that is implemented in that software package, resulting in the exclusion of 366 images because they were more than 2 SD away from the sample mean. After these quality control steps, images from 14,379 individuals were available for analysis. The vast

² CAT toolbox: www.neuro.uni-jena.de/cat/; SPM 12: www.fil.ion.ucl.ac.uk/spm/software/spm12/

majority of these 14,379 individuals reported to be of white European ancestry ($N=13,894$, UKB data field 21000).

Independent from us, the UKB Imaging Working Group also derived a measure of brain volume in a slightly smaller subsample ($N=14,165$) that is based on white and grey matter only (i.e. excluding fluid, see data field 25010 and (Miller et al. 2016)). The correlation between their measure of brain volume and our *TBV* is $r=.91$ ($p < 0.0001$). As a robustness check, we repeated our main analysis with the UKB-derived measure.

Genetic principal components

To control for ancestry and genetic diversity in the sample, we used the first 40 principal components (PCs) of the genetic data (for details, see (Bycroft et al. 2017)). The PCs were derived from high quality markers from all autosomes that were pruned to minimise linkage disequilibrium (Price et al. 2008), resulting in a set of 147,604 SNPs that were obtained from a set of 407,219 unrelated, high quality samples that matches our subsamples very closely in terms of ethnicity.

Descriptive statistics of the sample

Fig. S1 displays the distribution of *TBV* in our sample; the distributions of the cognitive scores and *EA* are displayed in **Fig. S2-S7**. The descriptive statistics of our sample are reported in **Table S3**, and **Table S2** summarizes the first order pairwise correlations between the key variables used in our analyses.

Among the different cognitive measures, fluid intelligence was most strongly correlated with our general intelligence factor (g), as well as *EA* and *TBV*. Male sex and body height had strong positive correlations with *TBV* and weak positive correlations with cognitive performance in the UKB sample. These findings highlight the importance of controlling for sex and height in our analyses.

We also observe small correlations ($|r| < .13$) between the first and second principal components of the genetic data with *TBV* and measures of cognitive performance, most noticeably for *fluid intelligence*. The first few genetic PCs in European samples typically map the settlement and historical migration patterns in a country relatively well. Thus, genetic PCs tend to capture environmental differences in terms of living standards, religion, and culture across people which may bias the estimated relationship between *TBV* and fluid intelligence if they are not controlled for.

Analysis

Our analyses followed a pre-registered protocol (<https://goo.gl/NJpUH3>). Specifically, we used UKB data from all European-descent individuals who were genotyped and scanned by April 2018 who also had measures of *fluid intelligence*, *EA* and all other control variables described in the protocol ($N=13,608$). We tested for an association between *TBV* (= white matter + grey matter + fluid) with *fluid intelligence* or *EA* using linear regression models that controlled for sex, age at brain scan, age at IQ test (using a dummy for each year to capture non-linear effects), all interactions between age at IQ test and sex, height, the indicator variables for the instances of the cognitive test, and the first 40 principal components of the genetic data.

For individuals who participated in more than one instance of the cognitive test, we computed and controlled for the average age at testing, rounded to the next integer value. The regressions on *EA* controlled for birth-year dummies instead of age at IQ measurement, to capture differences due to time-specific environmental factors (e.g., educational reforms). To estimate the marginal R^2 of *TBV* on *fluid intelligence* and *EA*, we computed the ΔR^2 between a model that includes all covariates (including genetic PCs) but no *TBV*, with a model that does include it.

In order to observe whether the relationship between *TBV* and cognitive performance is biased by subtle population structure and body height, we estimated additional models that do not include genetic controls or body height, and compared the coefficients with the model that included them. We further performed multiple regression analyses that decomposed the effect of *TBV* into grey and white matter, as well as fluid volume.

Our large sample also allowed us to conduct stratified analyses that elucidate whether the relationship between brain size and cognitive measures is constant across different population groups. Our analysis plan specified that subsamples need to be large enough to yield at least 90% statistical power to test effect sizes of $r > .1$ at $\alpha=0.05$ after Bonferroni correction for multiple comparison. Assuming we would conduct at most 50 independent tests ($\alpha=.05/50=.001$), the minimum required subsample size to achieve 90% power for an effect of $r=.1$ is $N=2,096$. Given this threshold, we were well-powered to conduct separate analyses for males ($N=6,425$), females ($N=7,183$), and four age groups, dividing the sample at the 25th, 50th, and 75th percentile of the age distribution ($N > 3,278$ in each group).

Our analysis plan also considered the possibility to compare effect sizes across groups of

different ancestry (e.g. European, Chinese, Indian). However, the vast majority of our final sample was of white European descent ($N=13,180$) and no other ethnic group was large enough to be studied separately given our predefined criteria for statistical power.

Apart from our pre-registered plan, we performed additional robustness checks by repeating the main analyses while replacing the dependent variables by the three additional cognitive tests available in the UKB (numeric memory, reaction time, and visual memory), as well as the four different proxies of g that we constructed. Furthermore, we ran regressions that added controls for place of birth (using dummy variables for geographic East / North coordinates) and socio-economic status, approximated by the Townsend deprivation index. The Townsend index is based on the postal code of a participant's household address and measures unemployment, non-car ownership, non-house ownership, and overcrowding in an area. Higher Townsend scores indicate higher deprivation (Hill et al. 2016).

Finally, we tested whether the association between *TBV* and cognitive performance is driven by a specific cognitive construct, by estimating a multiple linear regression model that predicts *TBV* from all four different cognitive tests and control variables.

Results

TBV and fluid intelligence

Fig. 1 illustrates the positive relationship between *TBV* and *fluid intelligence* in our pooled sample of $N=13,608$. We find a correlation between *TBV* and *fluid intelligence* of $r=.21$ (95%

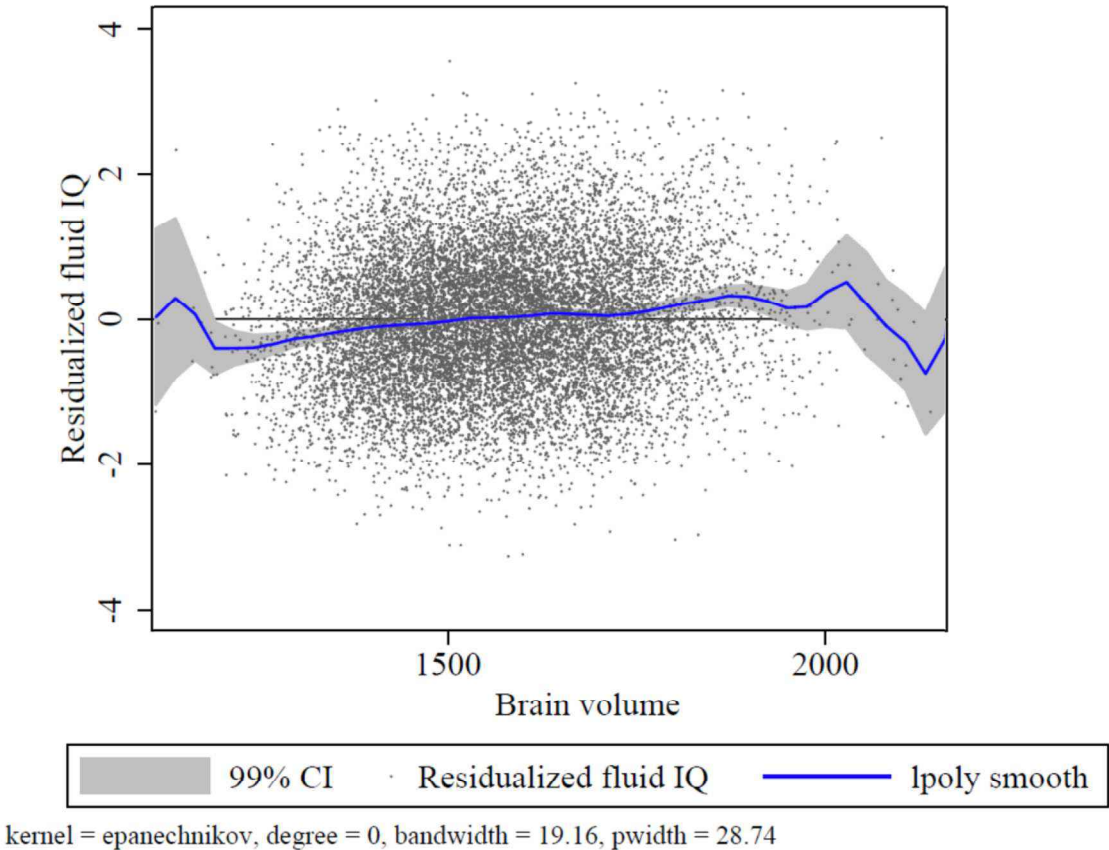
confidence interval, hereinafter 95% CI: .19 - .23, $p=3.20\times 10^{-86}$) without genetic controls, and $r=.19$ (95% CI: .17 - .22, $p=4.30\times 10^{-74}$) after correcting for subtle population structure (**Table 1**).³ Using the Townsend index of social deprivation and place of birth⁴ instead of genetic PCs yields exactly the same result ($r=.19$, 95% CI: .17 - .22, $N=12,822$, **Table S5**). Adding the genetic PCs to the regression that already controlled for the Townsend index and place of birth does not attenuate the association between brain volume and fluid intelligence any further. Thus, the relationship between *TBV* and *fluid intelligence* survives stringent controls for possible confounds. Without controlling for body height, the estimated relationship between *TBV* and *fluid intelligence* slightly increases to $r=.21$ (95% CI: .19 - .23, $p=2.52\times 10^{-92}$) (**Table S6**).

³ Similar results (i.e., significant coefficients for *TBV* and substantial overlap in the 95% CIs) are obtained when repeating the analyses for each of the test taking instances in isolation, in the sub-sample that took all four tests (see **Table S4**, $N=708$).

⁴ We report regression results with dummy variables for north / east coordinates; the results hold when dummy variables for all interactions of north / east coordinates are used instead.

Figure 1: Brain size and fluid intelligence

The relationship between total brain volume (*TBV*) and fluid intelligence, represented by a local polynomial smooth with 99% confidence intervals (in grey). Fluid intelligence was first normalized and then residualized by sex, age, height, the first 40 principal components of the genome, sex×age interactions and indicator variables for the instances of the cognitive tests taken as independent variables.



Overall, variation in *TBV* accounts for $\Delta R^2 \approx 2.1\%$ of the variation in *fluid intelligence* in the sample. The estimated marginal effects in the model including all controls suggest that a 100cm^3 increase in *TBV* at the population mean increases the expected *fluid intelligence* by .14 standard deviations (with sample $SD=1$, 95% CI: .13 - .16). Using the UKB-derived measure of brain volume ($N=13,409$), we find estimates with overlapping 95% CIs: A correlation of $r=.18$ (95% CI: .16 - .20, $p=5.82 \times 10^{-68}$) in the model including all controls and a marginal effect of .17 for each 100cm^3 increase in total white and grey matter (95% CI: .15 - .18, $p=5.82 \times 10^{-68}$, **Table S7**).

Table 1: Brain volume and fluid intelligence

Ordinary least squares (OLS) regression with *fluid intelligence* as the dependent variable. Table reports 95% confidence intervals in parentheses. Total brain volume is in cm³, control variables include sex (baseline category is female), age at scan in years, average age at IQ testing sessions (dummy coded) and its interactions with sex, height in cm and participant specific IQ testing sessions (dummy coded). The two right columns also include controls for population structure using the first 40 principal components of the genome. Coefficients for genetic PCs, indicators for IQ test, and age×sex interactions are not displayed.

	Excluding genetic controls		Including genetic controls	
	Standardized betas	Marginal effects (dy/dx)	Standardized betas	Marginal effects (dy/dx)
Total brain volume	0.21*** (0.19 - 0.23)	0.0014*** (0.0013 - 0.0016)	0.19*** (0.17 - 0.22)	0.0013*** (0.0012 - 0.0015)
Male	0.08 (-1.05 - 1.22)	-1.09 (-3.71 - 1.54)	0.21 (-0.91 - 1.33)	-0.15 (-2.75 - 2.44)
Age at scan	0.26*** (0.17 - 0.34)	0.04*** (0.02 - 0.05)	0.22*** (0.14 - 0.31)	0.03*** (0.02 - 0.04)
Height	0.11*** (0.09 - 0.14)	0.01*** (0.01 - 0.01)	0.09*** (0.06 - 0.11)	0.01*** (0.01 - 0.01)
R^2	0.11	0.11	0.13	0.13
N	13,608	13,608	13,608	13,608

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Including controls for potential confounds, our effect size estimate is 20% to 35% smaller than in the recent meta-analyses by Pietschnig et al. (2015) ($r = .24$, 95% CI: .21 - .27, $N = 8,036$) and G. E. Gignac and Bates (2017) ($r = .29$, 95% CI: .24 - .33). One potential reason is that we used more stringent controls for potential confounds than previous work. However, even the raw correlation between *TBV* and *fluid intelligence* in our data ($r = .20$) is smaller than in previous work. A likely cause underlying this smaller estimate is that *fluid intelligence* is measured with more noise in our data compared to other studies that used longer, more comprehensive cognitive tests (G. E. Gignac and Bates 2017). One way to account for measurement error is to divide the correlation between *fluid intelligence* and *TBV* by the test-retest reliability of the *fluid intelligence* measure,

which is between .60 and .69 (see **Table S1**).⁵ This leads to disattenuated effects of up to $r=.35$ (without genetic controls) and $r=.32$ (with controls), which is consistent with the estimates in the most recent meta-analyses of the literature (Pietschnig et al. 2015; G. E. Gignac and Bates 2017).

TBV and educational attainment

We also find a robust empirical relationship between *TBV* and *EA* (**Table 2**). Although *EA* is measured almost without error (in contrast to *fluid intelligence*), the correlation with *EA* is smaller than for *fluid intelligence* ($r=.12$, 95% CI: .10 - .15 including genetic controls, $N=13,608$). We find an almost identical result when using the Townsend index of social deprivation and place of birth as control variables for population structure instead of genetic PCs ($r=.11$, 95% CI: .08 - .13, $N=12,822$, **Table S8**). Repeating the regressions with the UKB-derived measure of *TBV* yields results with overlapping 95% confidence intervals to the main analyses (**Table S7**). Overall, *TBV* accounts for $\Delta R^2 \approx 0.9\%$ of the sample variation in *EA*. To put this result in perspective, an increase of 100cm^3 in *TBV* at the population mean increases the expected schooling by .4 years.

Grey matter, white matter, and fluid

Table 3 shows the results of a multiple regression that decomposed the effect of *TBV* into grey and white matter, as well as fluid volume. The largest contribution to *fluid intelligence* comes from grey matter ($r=.13$, 95% CI: .10 - .16). White matter ($r=.06$, 95% CI: .03 - .09) and fluid are also associated ($r=.05$, 95% CI: .03 - .07) with *fluid intelligence*, but to a much smaller extent. For *EA* we find comparable effect sizes of grey matter ($r=.06$, 95% CI: .03 - .09) and fluid

⁵ This approach assumes that the measurement noise of *TBV* is negligible.

($r=.07$, 95% CI: .05 - .09), and an even smaller effect of white matter that is indistinguishable from zero ($r=.03$, 95% CI: 0 - .06).

Analyses stratified by sex and age

The relationship between *TBV* and *fluid intelligence* is of comparable magnitude for females ($r=.16$, 95% CI: .14 - .18; $dy/dx=.0013$, 95% CI: .0011 - .0015) and males ($r=.15$, 95% CI: .13 - .17; $dy/dx=.0011$, 95% CI: .0010 - .0013, (**Table S9**). Furthermore, we find no interaction between sex and *TBV* influences on *fluid intelligence* (**Table S10**).

The relationship between *TBV* and *fluid intelligence* also appears to be relatively stable across age (**Table S11**). Although the effect size decreases to .15 in the oldest cohort (≥ 62 years), the 95% CI (.10 - .19) is overlapping with that of the other three age groups.

Our results for *EA* show a similar pattern. We find similar effect sizes for females ($r=.11$, 95% CI: .08 - .13) and males ($r=.09$, 95% CI: .07 - .12) as well as no significant age-dependent variation in effect sizes (**Tables S12, S13**).

Table 2: Brain volume and educational attainment

Ordinary least squares (OLS) regression with educational attainment as the dependent variable. Table reports 95% confidence intervals in parentheses. Brain volume is in cm³, control variables include sex (baseline category is female), age at scan in years, birth year (dummy coded) and its interactions with sex, height (in cm). The two right columns also include controls for population structure using the first 40 principal components of the genome. Coefficients for genetic PCs, and age×sex interactions are not displayed.

	Excluding genetic controls		Including genetic controls	
	Standardized betas	Marginal effects (dy/dx)	Standardized betas	Marginal effects (dy/dx)
Total brain volume	0.11*** (0.09 - 0.14)	0.0037*** (0.0030 - 0.0044)	0.12*** (0.10 - 0.15)	0.0040*** (0.0033 - 0.0047)
Male	0.53 (-0.28 - 1.34)	-8.38 (-19.60 - 2.84)	0.49 (-0.31 - 1.29)	-7.40 (-18.53 - 3.72)
Age at scan	-0.00 (-0.12 - 0.12)	-0.00 (-0.08 - 0.08)	-0.03 (-0.15 - 0.09)	-0.02 (-0.10 - 0.06)
Height	0.06*** (0.03 - 0.08)	0.03*** (0.02 - 0.04)	0.06*** (0.03 - 0.08)	0.03*** (0.02 - 0.04)
R ²	0.03	0.03	0.05	0.05
N	13,608	13,608	13,608	13,608

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 3: White, grey, and fluid matter separately

Ordinary least squares (OLS) regression with fluid intelligence (two left columns) and educational attainment (two right columns) as the dependent variables. Table reports 95% confidence intervals in parentheses. Total grey matter, white matter and fluid volumes are in cm^3 . Regressions include controls for population structure using the first 40 principal components of the genome and all other control variables specified in Table 1 (for fluid intelligence) and Table 2 (for educational attainment). Coefficients for control variables are not displayed.

	Fluid intelligence		Educational attainment	
	Standardized betas	Marginal effects (dy/dx)	Standardized betas	Marginal effects (dy/dx)
Grey matter	0.13*** (0.10 - 0.16)	0.0021*** (0.0016 - 0.0026)	0.06*** (0.03 - 0.09)	0.0010*** (0.0004 - 0.0015)
White matter	0.06*** (0.03 - 0.09)	0.0010*** (0.0005 - 0.0015)	0.03 (-0.00 - 0.06)	0.0004 (-0.0001 - 0.0009)
Fluid	0.05*** (0.03 - 0.07)	0.0008*** (0.0004 - 0.0012)	0.07*** (0.05 - 0.09)	0.0011*** (0.0008 - 0.0015)
R^2	0.14	0.14	0.06	0.06
N	13,608	13,608	13,608	13,608

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Robustness checks

We repeated our analysis with more elaborate proxies of g (**Tables S14a-d**).

For our primary proxy of g , we find almost identical standardized effect size estimates as in our main analysis on *fluid intelligence* ($r = .18$, 95% CI: .15 - .21 including genetic controls, $N = 7,511$; **Table S14a**). The same holds for the proxy of g derived by Lyall et al. (2016) ($r = .18$, 95% CI: .09 - .26, $N = 1,017$, **Table S14b**). We find slightly higher standardized effect sizes when using factor analysis instead of principal component analysis to derive g ($r = .21$, 95% CI: .18 - .24 including genetic controls, $N = 7,511$ (**Table S14c**)). However, the 95% CI of the estimates are all overlapping with our results for *fluid intelligence*. These findings are confirmed when we estimate marginal effects instead of standardized betas.

When using the g measure constructed without *fluid intelligence*, the relation with *TBV* was substantially smaller ($r=.10$, 95% CI: .07 - .12 including genetic controls, $N=7,511$; **Table S14d**), suggesting that a large share of the association between *TBV* and cognitive ability is accounted for by *fluid intelligence*.

Specificity

In order to explore the associations between *TBV* and cognitive measures that are different from *fluid intelligence* and g , we conducted exploratory analyses using the three other cognitive tasks of the UKB (see **Table S15**).

We find statistically significant, yet much smaller in magnitude associations of *TBV* with *numeric memory* ($r=.11$, 95% CI: .08 - .14 including genetic controls, $N=7,722$) and *visual memory* ($r=-.05$, 95% CI: -.07 - -.03 including genetic controls, $N=13,292$), and no significant relationship with the *reaction time* task ($r=-.02$, 95% CI: -.04 - .00 including genetic controls, $N=13,292$).

Moreover, when predicting *TBV* using a multinomial regression that includes the four different cognitive measures in our data altogether, the coefficient of *fluid intelligence* is substantially larger than the coefficients of all other measures (see **Table S16**), suggesting that the association between *TBV* and cognitive ability is best captured by *fluid intelligence*. This finding is robust to controlling for *EA* in the regression. It is important to note, however, that the smaller association of *TBV* with *numeric memory* and *visual memory* is likely driven by the low quality of these measures (see **Table S1**).

Discussion

Our results indicate a robust positive relationship between *TBV* and intelligence that is similar across sex and various age strata. When we account for the relatively low reliability of the cognitive measures in the UKB, the estimated effect sizes are comparable to previous recent meta-analyses on this topic. Yet, *TBV* accounts for a relatively small share in overall variation in cognitive performance ($\Delta R^2 \approx 2\%$). Importantly, our results are free of publication bias, come from a sample that is $\approx 70\%$ larger than all previous investigations on this topic combined, and systematically control for important potential confounds.

Our analysis shows that the lion's share of the association between *TBV* and intelligence is explained by individual differences in grey matter volume. Furthermore, we document that *TBV* is also positively associated with *EA*, although the association is substantially smaller than for intelligence ($\Delta R^2 \approx 0.9\%$).

While our study demonstrates that the association between *TBV* and cognitive performance is solid, our work and the literature as a whole have limitations that provide avenues for further research. First, our results are based on a large population sample of adults and elderly that over-represented individuals of higher socio-economic status, and consists almost entirely of European descent individuals from the UK. The positive, linear relationship between *TBV* and *fluid intelligence* that we observed was driven by the large majority of individuals in that sample who had normal range brain volumes and measures of *fluid intelligence*. At the extreme ends of the distributions, the relationship between *TBV* and fluid intelligence seems to be weaker or even

non-existent (**Fig. 1**). It is reasonable to expect that the positive relationship we observed would not hold for people affected by chronic or degenerative neurological problems (e.g., dementia, Alzheimer's disease, Parkinson's disease) or other medical conditions that are known to be linked to abnormal brain development or physiology. Furthermore, the results may not generalize to children. While we have no reason to believe that the results depend on other characteristics of the participants, materials, or context, continuous exploration of the generalizability of the results to other populations is worthwhile.

A second important limitation concerns causal inference. The empirical work on the relationship between *TBV* and intelligence or EA, including our study, is based on non-experimental data that cannot rule out reverse causation or the influence of unobserved confounds. Although it may be most intuitive that brain anatomy causes cognitive performance and EA, a reverse relationship may also exist (e.g. via brain plasticity that adapts the brain to how it is used, e.g., May 2011). Furthermore, although we control for more potential confounding factors than earlier studies, the identifying assumption of regression analysis that the error term is independent from the regressors may still be violated. For example, people with larger brains may have access to better schools and healthcare systems in a manner that is not captured by our genetic and demographic controls. In addition, brain anatomy and cognitive performance are both highly heritable ($h^2 \approx .8$, Posthuma et al. 2002), and the co-heritability between the two ($r_g \approx .3$, Sniekers et al. 2017) suggests that both are partially influenced by the same genetic factors (Posthuma et al. 2002; Okbay, Beauchamp, et al. 2016). Investigating these relationships further would be of interest.

Third, the low measurement quality of behavioral phenotypes in large datasets is a limitation that is the result of a trade-off between sample size and measurement accuracy, both of which are costly. While using a crude measure of a construct in a very large sample often allows obtaining greater statistical power than a perfect measure in a small sample (Okbay, Baselmans, et al. 2016), measurement error leads to attenuated (standardized) effect size estimates. We addressed this challenge by reporting disattenuated effects that divided sample estimates by the retest reliability of the cognitive measures.

Fourth, it is likely that structural differences in specific brain regions differentially contribute to individual differences in cognitive performance, over and above what is captured by *TBV*. Of note, despite a strong correlation between sex and *TBV* in our sample ($r=.62$), all of the cognitive measures in our sample showed sex differences that were meager (**Table S1**), suggesting the possibility that sex differences in other brain characteristics compensate for the discrepancy in *TBV* (e.g., females have greater cortical thickness; Ritchie et al. 2017).

Fifth, the relationship between anatomical brain features and cognitive performance is likely mediated by neural processes that are better captured by measures of functional brain activity rather than volumetric measurements. Furthermore, many distinct mental processes (e.g., attention and memory) contribute to performance in intelligence tests. Therefore, our understanding of how individual differences in cognition arise may benefit greatly from more detailed, possibly non-linear, mappings between anatomical and functional brain measures and individual differences in distinct mental capacities.

Finally, further theoretical accounts for what the association between *TBV* and intelligence might imply about the evolution of human intelligence are needed (e.g., González-Forero and Gardner 2018). Many previous investigations have been motivated by an implicit assumption that humans have particularly large brains and are also exceptionally cognitively flexible, relative to other species (Gonda, Herczeg, and Merilä 2013). However, there are no agreeable means to quantify intelligence between species, and although some recent efforts reported cross-species correlations between *TBV* and cognitive traits such as self-control (MacLean et al. 2014) and problem-solving (Benson-Amram et al. 2016), this emerging literature is in its early days, and is not without controversies (Kabadayi et al. 2016). Furthermore, humans are by no means the species with the largest brain size (cetaceans and elephants have much larger brains), brain to body size ratio, or relative number of neurons, and empirical evidence suggests that our species is also not superior when it comes to various cognitive phenotypes, including working memory (Inoue and Matsuzawa 2007).

We hope that future studies will shed further light on how individual differences in cognitive capacities arise by exploring the associations between cognitive abilities and additional biomarkers (such as functional brain measures), as well as their interactions with environmental conditions.

Acknowledgements

The research has been conducted using the UK Biobank Resource under application numbers 11425. Philipp Koellinger acknowledges financial support from an ERC Consolidator Grant (647648 EdGe). Gideon Nave acknowledges the financial support of the Wharton Neuroscience Initiative and Wharton’s Dean Research fund.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Authors contributions

G.N and P.D.K developed the study concept and design, performed data analysis and interpretation, and wrote the manuscript. W.J and R.K.L pre-processed the brain imaging data. All authors provided comments and approved the final version of the manuscript for submission.

References

- Batty, G. David, Karin Modig Wennerstad, George Davey Smith, David Gunnell, Ian J. Deary, Per Tynelius, and Finn Rasmussen. 2009. "IQ in Early Adulthood and Mortality by Middle Age: Cohort Study of 1 Million Swedish Men." *Epidemiology* 20 (1): 100–109.
- Benson-Amram, Sarah, Ben Dantzer, Gregory Stricker, Eli M. Swanson, and Kay E. Holekamp. 2016. "Brain Size Predicts Problem-Solving Ability in Mammalian Carnivores." *Proceedings of the National Academy of Sciences of the United States of America* 113 (9): 2532–37.
- Benyamin, B., Bst Pourcain, O. S. Davis, G. Davies, N. K. Hansell, M. J. Brion, R. M. Kirkpatrick, S. F. RAM Cents, M. B. Miller, and C. M. A. Haworth. 2013. "Childhood Intelligence Is Heritable, Highly Polygenic and Associated with FBNP1L." *Molecular Psychiatry* 19: 1–6.
- Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, et al. 2017. "Genome-Wide Genetic Data on ~500,000 UK Biobank Participants." *bioRxiv*. <https://doi.org/10.1101/166298>.
- Cardon, Lon R., and Lyle J. Palmer. 2003. "Population Stratification and Spurious Allelic Association." *The Lancet* 361 (9357): 598–604.
- Coffey, C. E., J. A. Saxton, G. Ratcliff, R. N. Bryan, and J. F. Lucke. 1999. "Relation of Education to Brain Size in Normal Aging: Implications for the Reserve Hypothesis." *Neurology* 53 (1): 189–96.
- Deary, Ian J., Steve Strand, Pauline Smith, and Cres Fernandes. 2007. "Intelligence and Educational Achievement." *Intelligence* 35 (1): 13–21.
- Deary, Ian J., Lawrence J. Whalley, Helen Lemmon, J. R. Crawford, and John M. Starr. 2000. "The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey." *Intelligence* 28 (1): 49–55.
- Francis Galton. 1889. "On Head Growth in Students at the University of Cambridge." *The Journal of the Anthropological Institute of Great Britain and Ireland* 18: 155–56.
- Gaser, Christian, and Robert Dahnke. 2016. "CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data." *HBM* 2016, 336–48.
- Gignac, Gilles E., and Timothy C. Bates. 2017. "Brain Volume and Intelligence: The Moderating Role of Intelligence Measurement Quality." *Intelligence* 64 (September): 18–29.
- Gignac, Gilles, Philip A. Vernon, and John C. Wickett. 2003. "Factors Influencing the Relationship between Brain Size and Intelligence." *The Scientific Study of General*

- Intelligence: Tribute to Arthur R. Jensen*, 93–106.
- Gonda, Abigél, Gábor Herczeg, and Juha Merilä. 2013. “Evolutionary Ecology of Intraspecific Brain Size Variation: A Review.” *Ecology and Evolution* 3 (8): 2751–64.
- Gonzales, J. E., and C. A. Cunningham. 2015. “The Promise of Pre-Registration in Psychological Research.” *Psychological Science Agenda*. Retrieved from [Http://www. Apa. org/science/about/psa/2015/08/pre-Registration](http://www.apa.org/science/about/psa/2015/08/pre-Registration). *Aspx*.
- González-Forero, Mauricio, and Andy Gardner. 2018. “Inference of Ecological and Social Drivers of Human Brain-Size Evolution.” *Nature* 557 (7706): 554–57.
- Greene, William H. 2003. *Econometric Analysis*. Edited by Pearson Education. Vol. 97. Prentice Hall.
- Haier, Richard J., Rex E. Jung, Ronald A. Yeo, Kevin Head, and Michael T. Alkire. 2004. “Structural Brain Variation and General Intelligence.” *NeuroImage* 23 (1): 425–33.
- Hill, W. David, Saskia P. Hagenaars, Riccardo E. Marioni, Sarah E. Harris, David C. M. Liewald, Gail Davies, Aysu Okbay, Andrew M. McIntosh, Catharine R. Gale, and Ian J. Deary. 2016. “Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank.” *Current Biology: CB* 26 (22): 3083–89.
- Inoue, Sana, and Tetsuro Matsuzawa. 2007. “Working Memory of Numerals in Chimpanzees.” *Current Biology: CB* 17 (23): R1004–5.
- Ioannidis, John P. A., Marcus R. Munafò, Paolo Fusar-Poli, Brian A. Nosek, and Sean P. David. 2014. “Publication and Other Reporting Biases in Cognitive Sciences: Detection, Prevalence, and Prevention.” *Trends in Cognitive Sciences* 18 (5): 235–41.
- Kabadayi, Can, Lucy A. Taylor, Auguste M. P. von Bayern, and Mathias Osvath. 2016. “Ravens, New Caledonian Crows and Jackdaws Parallel Great Apes in Motor Self-Regulation despite Smaller Brains.” *Royal Society Open Science* 3 (4): 160104.
- Lager, Anton Carl Jonas, and Jenny Torssander. 2012. “Causal Effect of Education on Mortality in a Quasi-Experiment on 1.2 Million Swedes.” *Proc. Natl. Acad. Sci.* 109 (22): 8461–66.
- Lyall, Donald M., Breda Cullen, Mike Allerhand, Daniel J. Smith, Daniel Mackay, Jonathan Evans, Jana Anderson, et al. 2016. “Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants.” *PloS One* 11 (4): e0154222.
- MacLean, Evan L., Brian Hare, Charles L. Nunn, Elsa Addessi, Federica Amici, Rindy C. Anderson, Filippo Aureli, et al. 2014. “The Evolution of Self-Control.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (20): E2140–48.
- May, Arne. 2011. “Experience-Dependent Structural Plasticity in the Adult Human Brain.” *Trends in Cognitive Sciences* 15 (10): 475–82.
- McDaniel, Michael A. 2005/7. “Big-Brained People Are Smarter: A Meta-Analysis of the Relationship between in Vivo Brain Volume and Intelligence.” *Intelligence* 33 (4): 337–46.
- Miller, Karla L., Fidel Alfaro-Almagro, Neal K. Bangerter, David L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, et al. 2016. “Multimodal Population Brain Imaging in the UK Biobank Prospective Epidemiological Study.” *Nature Neuroscience*, September. <https://doi.org/10.1038/nn.4393>.
- Okbay, Aysu, Bart M. L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G. Nivard, Mark Alan Fontana, S. Fleur W. Meddens, et al. 2016. “Genetic Variants Associated with Subjective Well-Being, Depressive Symptoms, and Neuroticism Identified through Genome-Wide Analyses.” *Nature Genetics* 48 (6): 624–33.
- Okbay, Aysu, Jonathan P. Beauchamp, Mark A. Fontana, James J. Lee, Tune H. Pers, Cornelius

- A. Rietveld, Patrick Turley, et al. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533: 539–42.
- Petersen, Steffen E., Paul M. Matthews, Fabian Bamberg, David A. Bluemke, Jane M. Francis, Matthias G. Friedrich, Paul Leeson, et al. 2013. "Imaging in Population Science: Cardiovascular Magnetic Resonance in 100,000 Participants of UK Biobank - Rationale, Challenges and Approaches." *Journal of Cardiovascular Magnetic Resonance: Official Journal of the Society for Cardiovascular Magnetic Resonance* 15 (May): 46.
- Pietschnig, Jakob, Lars Penke, Jelte M. Wicherts, Michael Zeiler, and Martin Voracek. 2015. "Meta-Analysis of Associations between Human Brain Volume and Intelligence Differences: How Strong Are They and What Do They Mean?" *Neuroscience and Biobehavioral Reviews* 57 (October): 411–32.
- Posthuma, Daniëlle, Eco J. C. De Geus, Wim F. C. Baaré, H. E. Hulshoff Pol, René S. Kahn, and Dorret I. Boomsma. 2002. "The Association between Brain Volume and Intelligence Is of Genetic Origin." *Nature Neuroscience* 5 (2): 83–84.
- Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38 (8): 904–9.
- Price, Alkes L., Michael E. Weale, Nick Patterson, Simon R. Myers, Anna C. Need, Kevin V. Shianna, Dongliang Ge, et al. 2008. "Long-Range LD Can Confound Genome Scans in Admixed Populations." *American Journal of Human Genetics* 83 (1): 132–35; author reply 135–39.
- Rietveld, Cornelius A., Dalton Conley, Nicholas Eriksson, Tõnu Esko, Sarah E. Medland, Anna A. E. Vinkhuyzen, Jian Yang, et al. 2014. "Replicability and Robustness of Genome-Wide Association Studies for Behavioral Traits." *Psychological Science* 25 (11): 1975–86.
- Rietveld, Cornelius A., Tõnu Tonu Esko, Gail Davies, Tune H. Pers, Patrick A. Turley, Beben Benyamin, Christopher F. Chabris, et al. 2014. "Common Genetic Variants Associated with Cognitive Performance Identified Using the Proxy-Phenotype Method." *Proceedings of the National Academy of Sciences of the United States of America* 111 (38): 13790–94.
- Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W. Martin, Harm-Jan Westra, et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340 (6139): 1467–71.
- Ritchie, S. J., S. R. Cox, X. Shen, M. V. Lombardo, and L. M. Reus. 2017. "Sex Differences in the Adult Human Brain: Evidence from 5,216 UK Biobank Participants." *BioRxiv*. <https://www.biorxiv.org/content/early/2017/04/14/123729.abstract>.
- Rubin, Donald B. 2004. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.
- Sniekers, Suzanne, Sven Stringer, Kyoko Watanabe, Philip R. Jansen, Jonathan R. I. Coleman, Eva Krapohl, Erdogan Taskesen, et al. 2017. "Genome-Wide Association Meta-Analysis of 78,308 Individuals Identifies New Loci and Genes Influencing Human Intelligence." *Nature Genetics*, May. <https://doi.org/10.1038/ng.3869>.
- Stephen Smith, Fidel Alfaro-Almagro And Karla Miller. 2014. "UK Biobank Brain Imaging Documentation." Wellcome Centre for Integrative Neuroimaging (WIN-FMRIB), Oxford University on behalf of UK Biobank. https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf.
- Stott, Denis H. 1983. "Brain Size and 'intelligence.'" *The British Journal of Developmental Psychology* 1 (3): 279–87.
- Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul

Downey, et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12 (3): e1001779.

Willerman, Lee, Robert Schultz, J. Neal Rutledge, and Erin D. Bigler. 1991/4. "In Vivo Brain Size and Intelligence." *Intelligence* 15 (2): 223–28.

For Review Only