

# Are Bigger Brains Smarter? Evidence From a Large-Scale Preregistered Study



Gideon Nave<sup>1</sup>, Wi Hoon Jung<sup>2,3</sup>, Richard Karlsson Linnér<sup>4</sup>,  
Joseph W. Kable<sup>1,3</sup>, and Philipp D. Koellinger<sup>4</sup>

<sup>1</sup>Marketing Department, The Wharton School, University of Pennsylvania; <sup>2</sup>Department of Psychology, Korea University; <sup>3</sup>Department of Psychology, University of Pennsylvania; and <sup>4</sup>Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam

Psychological Science  
 2019, Vol. 30(1) 43–54  
 © The Author(s) 2018  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)  
 DOI: 10.1177/0956797618808470  
[www.psychologicalscience.org/PS](http://www.psychologicalscience.org/PS)



## 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 controversial because of concerns about publication bias and the lack of systematic control for critical confounding factors (e.g., height, population structure). We conducted a preregistered study of the relationship between brain volume and cognitive performance using a new sample of adults from the United Kingdom that is about 70% larger than the combined samples of all previous investigations on this subject ( $N = 13,608$ ). Our analyses systematically controlled for sex, age, height, socioeconomic status, and population structure, and our analyses were free of publication bias. We found a robust association between total brain volume and fluid intelligence ( $r = .19$ ), which is consistent with previous findings in the literature after controlling for measurement quality of intelligence in our data. We also found a positive relationship between total brain volume and educational attainment ( $r = .12$ ). These relationships were mainly driven by gray matter (rather than white matter or fluid volume), and effect sizes were similar for both sexes and across age groups.

## Keywords

intelligence, educational attainment, brain volume, preregistered analysis, UK Biobank, open data, open materials, preregistered

Received 11/2/17; Revision accepted 8/9/18

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, are stable across the life span (Deary, Whalley, Lemmon, Crawford, & Starr, 2000), and are associated with important life outcomes, including educational attainment (Deary, Strand, Smith, & Fernandes, 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., Galton, 1889). For decades, the only way to test this

hypothesis was empirical studies using proxies of TBV such as head circumference. However, this work was controversial because of methodological issues (Stott, 1983) and concerns about racial and cultural bias.

The introduction of 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 ( $r$ ) of .51 in a sample of 40 college students (Willerman, Schultz, Neal Rutledge, & Bigler, 1991). However, the reported association has declined as sample sizes have grown: The first meta-analysis of the literature ( $k = 14$ ,  $N = 858$ ) estimated an

## Corresponding Author:

Philipp D. Koellinger, Vrije Universiteit Amsterdam, School of Business and Economics, Department of Economics, 1081 HV Amsterdam, The Netherlands  
 E-mail: [p.d.koellinger@vu.nl](mailto:p.d.koellinger@vu.nl)

average correlation of .37 (Gignac, Vernon, & Wickett, 2003). A later, more comprehensive meta-analysis ( $k = 37$ ,  $N = 1,530$ ) estimated a smaller correlation of .29 (McDaniel, 2005). The largest meta-analysis to date, which included unpublished data, reported an even smaller correlation of .24 ( $k = 88$ ,  $N = 8,036$ ; Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 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, researchers in only a few studies systematically controlled for confounding factors such as height, age, and socioeconomic status. A second concern is population stratification, that is, systematic biological differences across groups that might correlate with environmental and cultural factors.<sup>1</sup> If not properly controlled for, population stratification can induce a spurious relationship between biomarkers and phenotypes (Cardon & Palmer, 2003). For example, individuals of northwest 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 health care) 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).

A third issue is a bias toward publication of positive, statistically significant results and effect sizes that overestimate the true values. The most recent meta-analysis on intelligence and TBV by Pietschnig et al. (2015) found evidence for publication bias and showed that the correlation in published reports was .30 ( $k = 53$ ,  $N = 3,956$ ) but was only .17 in a larger set of unpublished studies ( $k = 67$ ,  $N = 2,822$ ). In contrast, Gignac and Bates (2017) did not find evidence for publication bias. However, their analysis was restricted to published studies of healthy participants only. Although several analytical techniques have been proposed to detect such bias, their capacity to estimate the true effect size is controversial, and their power to reject the null hypothesis of no publication bias is low in small samples (Ioannidis, Munafò, Fusar-Poli, Nosek, & David, 2014). A clean approach to avoid publication bias is to conduct a well-powered study following a preregistered analysis plan (Gonzales & Cunningham, 2015).

We addressed these three shortcomings of the current literature here. Specifically, we conducted a preregistered analysis of the relationship between measures of cognitive performance and TBV using data from the

UK Biobank (UKB; Miller et al., 2016; Sudlow et al., 2015). The UKB is a data collection of unprecedented richness and scale that was not part of any previous study on the relationship between TBV and cognitive performance. Our final sample contained 13,608 genotyped individuals with anatomical MRI brain scans. The sample was an adult population (> 40 years old) of European descent, all of whom completed at least one test of cognitive performance. This sample is approximately 70% larger than the combined samples of all previous studies associating in vivo TBV and intelligence (Pietschnig et al., 2015); it permits novel ways to control for confounds and allows comparing effect sizes across various demographic groups.

Our investigation provided the opportunity for two additional contributions. First, we investigated the differential contributions of gray 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 gray- and white-matter volumes are genetically correlated with general intelligence (Sniekers et al., 2017) and are thought to contribute to the association on the basis of small-sample studies (e.g., Haier, Jung, Yeo, Head, & Alkire, 2004); understanding their differential contributions is essential for further theoretical development of accounts of the relationship between TBV and intelligence.

Second, we examined the association between TBV and educational attainment, an important real-life outcome that crucially impacts individuals' income, health, and longevity (Lager & Torssander, 2012). To date, this association has been investigated in only a few small-sample studies of elderly or clinical populations (e.g., Coffey, Saxton, Ratcliff, Bryan, & Lucke, 1999).

## Method

### *The UKB data*

The UKB (Miller et al., 2016; Sudlow et al., 2015) recruited 502,617 people between the ages of 40 and 69 years in 2006 through 2010 from the general population across the entire United Kingdom. Almost all participants (488,363) have been genotyped (Bycroft et al., 2018), 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. As of April 2018, 15,040 participants have already been scanned, and their T1 structural brain images have been processed by the UKB team (Smith, Alfaro-Almagro, & Miller, 2014) and converted from

Digital Imaging and Communications in Medicine (DICOM) to Neuroimaging Informatics Technology Initiative (NIfTI) format. Health outcomes are tracked over time for all participants by linking the UKB to official hospital records. The principle goal of the project is to use large-scale longitudinal data to better understand disease etiology and to develop predictive methods for early onset disease detection. An important by-product of the UKB project is the generation of an unprecedentedly large and rich data set to study behavioral phenotypes and their relation to the collected biological markers (e.g., genotypes, brain scans) and health outcomes (e.g., cognitive performance, subjective well-being, body mass index, diseases).

## Measures

**Fluid intelligence.** The UKB contains a short measure of verbal-numerical reasoning (referred to as the fluid-intelligence test) that consists of 13 multiple-choice questions (see the Supplemental Material available online) measuring the capacity to solve problems that require logic and reasoning ability, independently of acquired knowledge. Participants had 2 min to complete as many questions as possible from the test. The fluid-intelligence test score is the simple unweighted sum of the number of correct answers given to these 13 questions. Participants who do not answer all of the questions within the allotted 2-min limit are given a score of zero for each of the unattempted questions.

The fluid-intelligence test was administered on three occasions: (a) the initial assessment visit, (b) the first repeat assessment visit, and (c) the imaging visit (see below). The test was also 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 ( $N$ s between 989 and 7,584; see Table S1 in the Supplemental Material), 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. We had access to 14,021 participants 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 of fluid intelligence for each participant by taking the average of these standardized scores (in cases in which multiple observations were available for an individual) and standardizing 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. participants who have taken only one test), we generated indicator variables for each one of the tests (i.e., a variable equal to 1 if the participant took a specific instance of the test and equal to 0 otherwise, and likewise for the other test instances) and included them as control variables in the regression analyses.

**Other cognitive measures.** Apart from conducting 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 by Lyall et al. (2016).

Numeric memory was measured by a task that first showed participants a 2-digit number and asked them to recall that number after a short pause. The number of digits then increased by 1 digit until either an error was made or the maximum number of 12 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 card game Snap, and each participant's 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 three or six 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, and Gignac and Bates (2017) found substantially higher correlations between brain size and cognitive ability in studies with "excellent" measures of IQ than in studies with "good" or "fair" measures ( $r = .39$ , 95% confidence interval, or CI = [.32, .46];  $r = .32$ , 95% CI = [.16, .46]; and  $r = .21$ , 95% CI = [.14, .28], respectively). To check the robustness of our main results based on the crude fluid-intelligence test described above, we repeated our analysis using four more comprehensive measures of  $g$ . 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  $g$  (Benyamin et al., 2014; Lyall et al., 2016; Rietveld, Esko, et al., 2014), yielding 7,511 participants.

We found that fluid intelligence had the highest loading on  $g$  (.77 in Lyall et al., 2016, and .78 in our data; see Table S2 in the Supplemental Material), consistent with the findings of earlier studies. In our analyses, we chose to focus on fluid intelligence instead of  $g$  because (a) the numeric memory test was available for only a subset of our participants, which reduced the sample size for  $g$  analyses by almost 50% compared with the sample size for fluid intelligence, and (b) imputation of missing observations was not possible without potentially introducing substantial noise. 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 (reaction time  $r \approx .55$ , visual memory  $r \approx .21$ ; see Table S1).

Our second measure of  $g$  was constructed by performing factor analysis of a single factor on the four tests instead of principal component analysis (PCA). The analysis used minimum residuals estimation and oblimin rotation. The correlation between this measure of  $g$  and our primary measure derived from PCA was .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 made use of the data from only the first touch-screen interview; it ignored data from the three-pair version of the pair-matching test and used natural logarithm (LN) transformations of reaction time and LN + 1 of the visual memory tests. PCA was then used as a dimensionality-reduction technique to extract  $g$  ( $N = 1,017$ ).

Our fourth measure of  $g$  was constructed in a similar manner to our primary measure but excluded the fluid-intelligence scores before PCA was performed ( $N = 7,511$ ). This provided a measure of  $g$  that did not directly depend on our main fluid-intelligence measure.

These four measures of intelligence would be rated as “good” according to the guidelines established by Gignac and Bates (2017; i.e., two to eight tests, two to three dimensions, 20- to 39-min testing time), compared with a “poor” rating of our main measure of fluid intelligence (one test, one dimension, very short testing time). However, fluid intelligence allowed us to study the relationship with TBV in a substantially larger sample ( $N = 13,608$  compared with  $Ns = 1,017$ – $7,511$ ).

**Educational attainment.** Following the standard established by the Social Science Genetic Association Consortium (Rietveld et al., 2013), we measured educational attainment as equivalent to years of U.S. schooling for the highest educational degree that an individual obtained. We followed the International Standard Classification of

Education (United Nations Educational, Scientific and Cultural Organization, 1997), which leads to seven categories of educational attainment that are internationally comparable. Educational attainment was measured via self-reports in the UKB on three occasions: (a) the initial assessment visit, (b) the first repeat assessment visit, and (c) the imaging visit. We used the highest educational degree reported on any of these occasions as our measure of educational attainment.

**TBV.** 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) = 2,000 ms, echo time (TE) = 2.1 ms, flip angle = 8°, matrix size = 256 × 256 mm, voxel size = 1 mm × 1 mm × 1 mm, 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 (CAT; Version12; [www.neuro.uni-jena.de/cat/](http://www.neuro.uni-jena.de/cat/)) implemented in Statistical Parametric Mapping (SPM) software (Version 12; Wellcome Centre for Human Neuroimaging; [www.fil.ion.ucl.ac.uk/spm/software/spm12/](http://www.fil.ion.ucl.ac.uk/spm/software/spm12/)). CAT12 is a fully automated toolbox for measurements of gray-matter and white-matter 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 gray matter, white matter, and cerebrospinal fluid; spatially normalized to Montreal Neurological Institute space using linear and nonlinear transformations; and modulated to preserve the total amount of signal in the original image during spatial normalization. TBV was calculated by summing the raw volumes of gray matter, white matter, and cerebrospinal fluid.

We conducted the following checks to ensure quality. First, we visually inspected all T1 images that were available to us as of April 2018 ( $N = 14,793$ ) and excluded 48 images because of artifacts, poor image quality, or gross brain pathology hampering image segmentation. Next, we processed the images using the CAT12 toolkit (Gaser & Dahnke, 2016) and performed the sample homogeneity check implemented in that software package, resulting in the exclusion of 366 images because they were more than 2 standard deviations away from the sample mean. After these quality control steps were conducted, images from 14,379 individuals were available for analysis. The vast majority of these 14,379 individuals reported to be of White European ancestry ( $N = 13,894$ , field 21000 in the UKB data set).

Independently from us, the UKB Imaging Working Group also derived a measure of brain volume in a slightly smaller subsample ( $n = 14,165$ ) based on white and gray matter only (i.e., excluding fluid; see field 25010 in the UKB data set and Miller et al., 2016). The

correlation between their measure of brain volume and our TBV is .91 ( $p < .0001$ ). To check the robustness, 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 of the genetic data (for details, see Bycroft et al., 2018). The principal components were derived from high-quality markers from all autosomes that were pruned to minimize linkage disequilibrium (Price et al., 2008), resulting in a set of 147,604 single-nucleotide polymorphisms obtained from a set of 407,219 unrelated, high-quality samples that match our subsamples very closely in terms of ethnicity.

**Descriptive statistics of the sample.** Figure S1 in the Supplemental Material displays the distribution of TBV in our sample; the distributions of the cognitive scores and educational attainment are displayed in Figures S2 to S7 in the Supplemental Material. The descriptive statistics of our sample are reported in Table S3 in the Supplemental Material, 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  $g$  as well as educational attainment 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 observed small correlations ( $|r| < .13$ ) between (a) the first and second principal components of the genetic data and TBV and (b) the first and second principal components of the genetic data and measures of cognitive performance, most noticeably for fluid intelligence. The first few genetic principal components in European samples typically map the settlement and historical migration patterns in a country relatively well. Thus, genetic principal components 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 preregistered protocol (<https://osf.io/fvm7p/register/565fb3678c5e4a66b5582f67>). Specifically, we used UKB data from all individuals of European descent who were genotyped and scanned by April 2018 who also had measures of fluid intelligence, educational attainment, and all other control variables described in the protocol ( $N = 13,608$ ). We tested for

an association between TBV (white matter + gray matter + fluid) and fluid intelligence and between TBV and educational attainment using linear regression models that controlled for sex, age at brain scan, age at IQ test (using a dummy for each year to capture nonlinear effects), height, the indicator variables for the instances of the cognitive test, the first 40 principal components of the genetic data, and all interactions between age at IQ test and sex.

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 educational attainment 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 educational attainment, we computed the change in  $R^2$  between a model that includes all covariates (including genetic principal components) but no TBV and a model that did not include them.

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

Our large sample also allowed us to conduct stratified analyses that elucidated whether the relationship between brain size and cognitive measures was constant across different population groups. Our analysis plan specified that subsamples needed to be large enough to yield at least 90% statistical power to test effect sizes with a correlation greater than .10 and an alpha of .05 after Bonferroni correction for multiple comparisons. Assuming that we would conduct, at most, 50 independent tests ( $\alpha = .05/50 = .001$ ), the minimum required subsample size to achieve 90% power for an effect ( $r$ ) of .10 would be 2,096. Given this threshold, we were well powered to conduct separate analyses for men ( $n = 6,425$ ), women ( $n = 7,183$ ), and four age groups, dividing the sample at the 25th, 50th, and 75th percentiles of the age distribution ( $n > 3,278$  in each group).

Our analysis plan also considered the possibility of comparing 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 preregistered plan, we performed additional robustness checks by repeating the main analyses while replacing the dependent variables by

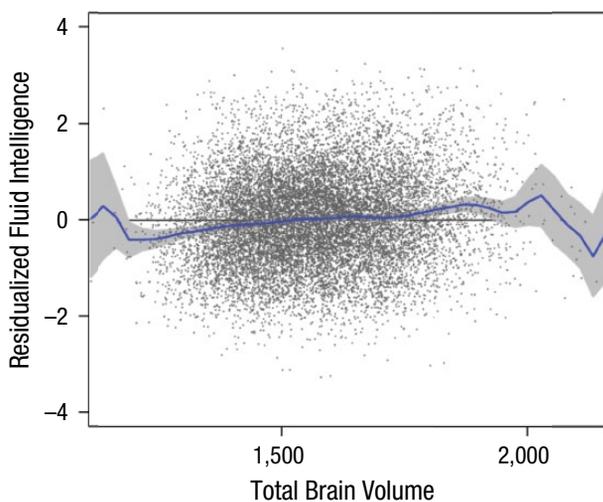
the three additional cognitive tests available in the UKB (numeric memory, reaction time, and visual memory), as well as the four different proxies of  $g$  that we constructed. Furthermore, we ran regressions that added controls for place of birth (using dummy variables for geographic east and north coordinates) and socioeconomic 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, lack of car ownership, lack of 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 was driven by a specific cognitive construct by estimating a multiple linear regression model that predicted TBV from all four different cognitive tests and control variables.

## Results

### *TBV and fluid intelligence*

Figure 1 illustrates the positive relationship between TBV and fluid intelligence in our pooled sample of 13,608 participants. We found a correlation between TBV and fluid intelligence of .21 (95% CI = [.19, .23],  $p = 3.20 \times 10^{-86}$ ) without genetic controls and .19 (95% CI = [.17, .22],  $p = 4.30 \times 10^{-74}$ ) after correcting for subtle population structure (see Table 1).<sup>2</sup> Using the Townsend index of social deprivation and place of birth<sup>3</sup> instead of genetic



**Fig. 1.** Scatterplot showing the relationship between total brain volume and residualized fluid intelligence. The regression line (in blue) was estimated using local polynomial smoothing. The gray band indicates 99% confidence intervals (CIs). Fluid intelligence was first normalized and then residualized by sex, age, height, the first 40 principal components of the genome, Sex  $\times$  Age interactions, and indicator variables for the instances of the cognitive tests taken as independent variables. Parameters for smoothing were as follows: kernel = epanechnikov, degree = 0, bandwidth = 19.16, pwidth = 28.74.

principal components yielded exactly the same result ( $r = .19$ , 95% CI = [.17, .22],  $N = 12,822$ ; see Table S5 in the Supplemental Material). Adding the genetic principal components to the regression that already controlled for the Townsend index and place of birth did not attenuate the association between brain volume and fluid intelligence any further. Thus, the relationship between TBV and fluid intelligence survived stringent controls for possible confounds. Without controlling for body height, we found that the estimated correlation between TBV and fluid intelligence slightly increased to .21 (95% CI = [.19, .23],  $p = 2.52 \times 10^{-92}$ ; see Table S6 in the Supplemental Material).

Overall, variation in TBV accounted for a change in  $R^2$  of approximately 2.1% of the variation in fluid intelligence in the sample. The estimated marginal effects in the model including all controls suggest that a 100-cm<sup>3</sup> increase in TBV at the population mean increased the expected fluid intelligence by 0.14 standard deviations (with sample  $SD = 1.0$ , 95% CI = [0.13, 0.16]). Using the UKB-derived measure of brain volume ( $N = 13,409$ ), we found estimates with overlapping 95% CIs: a correlation of .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 100-cm<sup>3</sup> increase in total white and gray matter (95% CI = [.15, .18],  $p = 5.82 \times 10^{-68}$ ; see Table S7 in the Supplemental Material).

When we included controls for potential confounds, our effect-size estimate was 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 Gignac and Bates (2017;  $r = .29$ , 95% CI = [.24, .33]). One potential reason is that we used more stringent controls for potential confounds than were used in 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 study compared with other studies, which used longer, more comprehensive cognitive tests (Gignac & Bates, 2017). One way to account for measurement error is to divide the correlation between fluid intelligence and TBV by the square root of the test-retest reliability of the fluid-intelligence measure, which is between .60 and .69 (see Table S1).<sup>4</sup> This leads to disattenuated effects ( $r_s$ ) of up to .27 (without genetic controls) and .25 (with controls), which are consistent with the estimates in the most recent meta-analyses in the literature (Gignac & Bates, 2017; Pietschnig et al., 2015).

### *TBV and educational attainment*

We also found a robust empirical relationship between TBV and educational attainment (see Table 2). Although educational attainment was measured almost without

**Table 1.** Results From the Ordinary Least Squares Regressions Testing the Influence of Total Brain Volume on Fluid Intelligence

Variable	Excluding genetic controls		Including genetic controls	
	$\beta$	Marginal effect ( $dy/dx$ )	$\beta$	Marginal effect ( $dy/dx$ )
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]
Control: 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]
Control: 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]
Control: 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]
$R^2$	.11	.11	.13	.13
$N$	13,608	13,608	13,608	13,608

Note: Values in brackets are 95% confidence intervals. Total brain volume was measured in cubic centimeters; control variables included sex (baseline category was female), age at scan in years, height in centimeters, participant-specific IQ testing sessions (dummy coded), and all interactions between average age at IQ-testing sessions (dummy coded) and sex. The two right columns also include controls for population structure using the first 40 principal components of the genome. Coefficients for genetic principal components, indicators for IQ test, and Age  $\times$  Sex interactions are not displayed. \* $p < .001$ .

error (in contrast to fluid intelligence), the correlation with educational attainment was smaller than for fluid intelligence ( $r = .12$ , 95% CI = [.10, .15] including genetic controls,  $N = 13,608$ ). We found 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 principal components ( $r = .11$ , 95% CI = [.08, .13],  $N = 12,822$ ; see Table S8 in the Supplemental Material). Repeating the regressions with the UKB-derived measure of TBV yielded results with 95% CIs that overlapped with the main analyses (see Table S7). Overall, TBV accounts for a change in  $R^2$  of approximately 0.9% of the sample variation in

educational attainment. To put this result in perspective, an increase of 100 cm<sup>3</sup> in TBV at the population mean increased the expected schooling by 0.4 years.

### ***Gray-matter, white-matter, and fluid volume***

Table 3 shows the results of a multiple regression that decomposed the effect of TBV into gray- and white-matter as well as fluid volume. The largest contribution to fluid intelligence came from gray matter ( $r = .13$ , 95% CI = [.10, .16]). White matter ( $r = .06$ , 95% CI = [.03, .09]) and fluid were also associated ( $r = .05$ , 95% CI =

**Table 2.** Results From the Ordinary Least Squares Regressions Testing the Influence of Total Brain Volume on Educational Attainment

Variable	Excluding genetic controls		Including genetic controls	
	$\beta$	Marginal effect ( $dy/dx$ )	$\beta$	Marginal effect ( $dy/dx$ )
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]
Control: 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]
Control: 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]
Control: 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]
$R^2$	.03	.03	.05	.05
$N$	13,608	13,608	13,608	13,608

Note: Values in brackets are 95% confidence intervals. Brain volume was measured in cubic centimeters; control variables included sex (baseline category was female), age at scan in years, birth year (dummy coded) and its interactions with sex, and height in centimeters. The two right columns also include controls for population structure using the first 40 principal components of the genome. Coefficients for genetic principal components and Age  $\times$  Sex interactions are not displayed. \* $p < .001$ .

**Table 3.** Results From the Ordinary Least Squares Regressions Testing the Influence of White-Matter, Gray-Matter, and Fluid Volume on Fluid Intelligence and Educational Attainment

Variable	Fluid intelligence		Educational attainment	
	Standardized $\beta$	Marginal effect ( $dy/dx$ )	Standardized $\beta$	Marginal effect ( $dy/dx$ )
Gray-matter volume	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 volume	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 volume	0.05* [0.03, 0.07]	0.0008* [0.0004, 0.0012]	0.07* [0.05, 0.09]	0.0011* [0.0008, 0.0015]
$R^2$	.14	.14	.06	.06
$N$	13,608	13,608	13,608	13,608

Note: Values in brackets are 95% confidence intervals. Total gray-matter, white-matter, and fluid volumes were measured in cubic centimeters. Regressions included 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.

\* $p < .001$ .

[.03, .07]) with fluid intelligence but to a much smaller extent. For educational attainment, we found comparable effect sizes of gray matter ( $r = .06$ , 95% CI = [.03, .09]) and fluid ( $r = .07$ , 95% CI = [.05, .09]) and an even smaller effect of white matter that was indistinguishable from 0 ( $r = .03$ , 95% CI = [.00, .06]).

### Analyses stratified by sex and age

The relationship between TBV and fluid intelligence was of comparable magnitude for women ( $r = .16$ , 95% CI = [.14, .18];  $dy/dx = 0.0013$ , 95% CI = [0.0011, 0.0015]) and men ( $r = .15$ , 95% CI = [.13, .17];  $dy/dx = 0.0011$ , 95% CI = [0.0010, 0.0013]; see Table S9 in the Supplemental Material). Furthermore, we found no interaction between sex and TBV influences on fluid intelligence (see Table S10 in the Supplemental Material). The relationship between TBV and fluid intelligence also appears to be relatively stable across age (see Table S11 in the Supplemental Material). Although the effect size decreased to .15 in the oldest cohort ( $\geq 62$  years), the 95% CI (.10, .19) overlapped with that of the other three age groups.

Our results for educational attainment show a similar pattern. We found similar effect sizes for women ( $r = .11$ , 95% CI = [.08, .13]) and men ( $r = .09$ , 95% CI = [.07, .12]) as well as no significant age-dependent variation in effect sizes (see Tables S12 and S13 in the Supplemental Material).

### Robustness checks

We repeated our analysis with more elaborate proxies of  $g$  (see Tables S14a–S14d in the Supplemental Material). For our primary proxy of  $g$ , we found standardized

effect-size estimates almost identical to those in our main analysis on fluid intelligence ( $r = .18$ , 95% CI = [.15, .21] including genetic controls,  $N = 7,511$ ; see Table S14a). The same held for the proxy of  $g$  derived by Lyall et al. (2016;  $r = .18$ , 95% CI = [.09, .26],  $N = 1,017$ ; see Table S14b). We found slightly higher standardized effect sizes when using factor analysis instead of PCA to derive  $g$  ( $r = .21$ , 95% CI = [.18, .24] including genetic controls,  $N = 7,511$ ; see Table S14c). However, the 95% CIs of the estimates all overlapped with our results for fluid intelligence. These findings were confirmed when we estimated marginal effects instead of 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,  $N = 7,511$ ; see Table S14d), suggesting that a large share of the association between TBV and cognitive ability is accounted for by fluid intelligence.

### Specificity

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 in the Supplemental Material). We found 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 TBV was predicted using a multinomial regression that included the four different cognitive

measures in our data altogether, the coefficient of fluid intelligence was substantially larger than the coefficients of all other measures (see Table S16 in the Supplemental Material), suggesting that the association between TBV and cognitive ability is best captured by fluid intelligence. This finding was robust to controlling for educational attainment in the regression. It is important to note, however, that the smaller association of TBV with numeric memory and visual memory was likely driven by the low quality of these measures (see Table S1).

## Discussion

Our results indicate that there is a robust positive relationship between TBV and intelligence that is similar across sex and various age strata. When we accounted for the relatively low reliability of the cognitive measures in the UKB, the estimated effect sizes were comparable with 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 and come from a sample that is approximately 70% larger than the combined samples of all previous investigations on this topic, and our analyses systematically controlled 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 gray-matter volume. Furthermore, we document that TBV is also positively associated with educational attainment, although the association is substantially smaller than for intelligence ( $\Delta R^2 \approx 0.9\%$ ).

Although 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 the elderly that overrepresented individuals of higher socioeconomic status, and the sample consists almost entirely of individuals of European descent from the United Kingdom. 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 brain volumes and measures of fluid intelligence in the normal range. At the extreme ends of the distributions, the relationship between TBV and fluid intelligence seems to be weaker or even nonexistent (see 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.

Although 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 and between TBV and educational attainment, including our study, is based on nonexperimental data, so we cannot rule out reverse causation or the influence of unobserved confounds. Although it may be most intuitive that brain anatomy causes cognitive performance and educational attainment, 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 controlled for more potential confounding factors than did authors of 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 health-care 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 ( $b^2 \approx .8$ ; Posthuma et al., 2002), and the coheritability between the two ( $r_g \approx .3$ ; Sniekers et al., 2017) suggests that both are partially influenced by the same genetic factors (Okbay, Beauchamp, et al., 2016; Posthuma et al., 2002). Investigating these relationships further would be of interest.

Third, the low measurement quality of behavioral phenotypes in large data sets is a limitation that is the result of a trade-off between sample size and measurement accuracy, both of which are costly. Whereas 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 square root of 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 (see Table S1), suggesting the possibility that sex differences in other brain characteristics compensate for the discrepancy in TBV (e.g., women have greater cortical thickness; Ritchie et al., 2018).

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 than by 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 nonlinear, 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 & 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, & 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, Dantzer, Stricker, Swanson, & Holekamp, 2016), this emerging literature is in its early days and is not without controversies (Kabadayi, Taylor, von Bayern, & Osvath, 2016). Furthermore, humans are by no means the species with the largest brain size (cetaceans and elephants have much larger brains), ratio of brain to body size, 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 & 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.

### Action Editor

Ralph Adolphs served as action editor for this article.

### Author Contributions

G. Nave and P. D. Koellinger developed the study concept and design, analyzed and interpreted the data, and wrote the manuscript. W. H. Jung and R. Karlsson Linnér preprocessed the brain-imaging data. All the authors provided comments and approved the final manuscript for submission.

### Acknowledgments

The research was conducted using the UK Biobank resource under Application No. 11425.

### Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Funding

P. D. Koellinger acknowledges financial support from a European Research Council Consolidator Grant (647648 EdGe). G. Nave acknowledges the financial support of the Wharton Neuroscience Initiative and The Wharton School's Dean Research Fund.

### Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618808470>

### Open Practices



All data and materials are available via UK Biobank at <http://www.ukbiobank.ac.uk/>. Data scripts for the present analyses can be found on the Open Science Framework (OSF) at <https://osf.io/x8rnq/>. The design and analysis plans were preregistered on the OSF at <https://osf.io/fvm7p/register/565fb3678c5e4a66b5582f67>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618808470>. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

### Notes

1. Population stratification is a well-known concern in genetic-association studies. For example, a spurious relationship between the lactase gene that codes for the enzyme lactase and educational attainment is found if analyses in 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 southeastern parts of Europe than in northwestern parts.
2. Similar results (i.e., significant coefficients for TBV and substantial overlap in the 95% CIs) were obtained when we repeated the analyses for each of the test-taking instances in isolation, in the subsample that took all four tests ( $N = 708$ ; see Table S4 in the Supplemental Material).
3. We report regression results with dummy variables for east and north coordinates; the results held when dummy variables for all interactions of east and north coordinates were used instead.
4. This approach assumes that the measurement noise of TBV is negligible.

### References

- Batty, G. D., Wennerstad, K. M., Smith, G. D., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2009). IQ in

- early adulthood and mortality by middle age: Cohort study of 1 million Swedish men. *Epidemiology*, *20*, 100–109.
- Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences, USA*, *113*, 2532–2537.
- Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M. J., . . . Visscher, P. M. (2014). Childhood intelligence is heritable, highly polygenic and associated with *FBNP1L*. *Molecular Psychiatry*, *19*, 253–258.
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*, 203–209. doi:10.1038/s41586-018-0579-z
- Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*, 598–604.
- Coffey, C. E., Saxton, J. A., Ratcliff, G., Bryan, R. N., & Lucke, J. F. (1999). Relation of education to brain size in normal aging: Implications for the reserve hypothesis. *Neurology*, *53*, 189–196.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49–55.
- Galton, F. (1889). On head growth in students at the University of Cambridge. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *18*, 155–156.
- Gaser, C., & Dahnke, R. (2016). CAT—a Computational Anatomy Toolbox for the analysis of structural MRI data. *Human Brain Mapping*, *37*, 336–348.
- Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence*, *64*, 18–29.
- Gignac, G., Vernon, P. A., & Wickett, J. C. (2003). Factors influencing the relationship between brain size and intelligence. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 93–106). Amsterdam, The Netherlands: Elsevier.
- Gonda, A., Herczeg, G., & Merilä, J. (2013). Evolutionary ecology of intraspecific brain size variation: A review. *Ecology and Evolution*, *3*, 2751–2764.
- Gonzales, J. E., & Cunningham, C. A. (2015, August). The promise of pre-registration in psychological research. *Psychological Science Agenda*. Retrieved from <http://www.apa.org/science/about/psa/2015/08/pre-Registration.aspx>
- González-Forero, M., & Gardner, A. (2018). Inference of ecological and social drivers of human brain-size evolution. *Nature*, *557*, 554–557.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *NeuroImage*, *23*, 425–433.
- Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., . . . Deary, I. J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology*, *26*, 3083–3089.
- Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004–R1005.
- Ioannidis, J. P. A., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences*, *18*, 235–241.
- Kabadayi, C., Taylor, L. A., von Bayern, A. M. P., & Osvath, M. (2016). Ravens, New Caledonian crows and jackdaws parallel great apes in motor self-regulation despite smaller brains. *Royal Society Open Science*, *3*(4), Article 160104. doi:10.1098/rsos.160104
- Lager, A. C. J., & Torrsander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences, USA*, *109*, 8461–8466.
- Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., . . . Pell, J. P. (2016). Cognitive test scores in UK Biobank: Data reduction in 480,416 participants and longitudinal stability in 20,346 participants. *PLOS ONE*, *11*(4), Article e0154222. doi:10.1371/journal.pone.0154222
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., . . . Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences, USA*, *111*, E2140–E2148.
- May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, *15*, 475–482.
- McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, *33*, 337–346.
- Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., . . . Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*, 1523–1536. doi:10.1038/nn.4393
- Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., . . . Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*, 624–633.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539–542.
- Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (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*, 411–432.
- Posthuma, D., De Geus, E. J. C., Baaré, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, *5*, 83–84.

- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909.
- Price, A. L., Weale, M. E., Patterson, N., Myers, S. R., Need, A. C., Shianna, K. V., . . . Reich, D. (2008). Long-range LD can confound genome scans in admixed populations. *American Journal of Human Genetics*, *83*(1), 132–135; author reply 135–139.
- Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A. E., . . . Social Science Genetics Association Consortium. (2014). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological Science*, *25*, 1975–1986.
- Rietveld, C. A., Esko, T. T., Davies, G., Pers, T. H., Turley, P. A., Benjamin, B., . . . Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences, USA*, *111*, 13790–13794.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467–1471.
- Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., . . . Deary, I. J. (2018). Sex differences in the adult human brain: Evidence from 5216 UK Biobank participants. *Cerebral Cortex*, *28*, 2959–2975.
- Smith, S., Alfaro-Almagro, F., & Miller, K. (2014). *UK Biobank brain imaging documentation* (Version 1.4). Wellcome Centre for Integrative Neuroimaging, Oxford University. Retrieved from [https://biobank.ctsu.ox.ac.uk/crystal/docs/brain\\_mri.pdf](https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf)
- Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., . . . Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, *49*, 1107–1112. doi:10.1038/ng.3869
- Stott, D. H. (1983). Brain size and “intelligence.” *The British Journal of Developmental Psychology*, *1*, 279–287.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Collins, R. (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), Article e1001779. doi:10.1371/journal.pmed.1001779
- United Nations Educational, Scientific and Cultural Organization. (1997). *ISCED mappings*. Retrieved from <http://uis.unesco.org/en/isced-mappings>
- Willerman, L., Schultz, R., Neal Rutledge, J., & Bigler, E. D. (1991). *In vivo* brain size and intelligence. *Intelligence*, *15*, 223–228.