Sex differences in brain size and general intelligence (g)

Dimitri van der Linden[a,*], Curtis S. Dunkel[b], Guy Madison[c]

[a] Institute of Psychology, Erasmus University Rotterdam, The Netherlands
[b] Department of Psychology, Western Illinois University, USA
[c] Department of Psychology, Umeå University, Sweden

ARTICLE INFO

ABSTRACT

Utilizing MRI and cognitive tests data from the Human Connectome project ($N = 900$), sex differences in general intelligence (g) and molar brain characteristics were examined. Total brain volume, cortical surface area, and white and gray matter correlated 0.1–0.3 with g for both sexes, whereas cortical thickness and gray/white matter ratio showed less consistent associations with g. Males displayed higher scores on most of the brain characteristics, even after correcting for body size, and also scored approximately one fourth of a standard deviation higher on g. Mediation analyses and the Method of Correlated Vectors both indicated that the sex difference in g is mediated by general brain characteristics. Selecting a subsample of males and females who were matched on g further suggest that larger brains, on average, lead to higher g, whereas similar levels of g do not necessarily imply equal brain sizes.

1. Introduction

Characteristics of the brain, such as its size, the density of neurons, and the proportion of gray and white matter have been shown to relate to various cognitive abilities (Colom et al., 2009; Wickett, Vernon, & Lee, 1994). General intelligence is the measure of cognitive ability that has received the most attention in this context, in accordance with its status as representing the overall efficiency to process information and solve novel problems (Duncan, Seitz, Kolodny, et al., 2000; Jensen, 1998). A widely used operationalization of general intelligence is g; a latent factor representing the proportion of common variance across a wide range of cognitive tasks (Jensen, 1998). The g factor is an important predictor of many of real-life outcomes such as educational attainment, job performance and health (e.g., Gottfredson, 1997).

Several meta-analyses have provided reliable estimates of the associations between g and brain characteristics. For example, McDaniel (2005) reported a population meta-analytic correlation between g and brain size of 0.33. Recently, Pietschnig, Penke, Wicherts, Zeiler, and Voracek (2015) expanded the number of studies, and arrived at a revised meta-analytic correlation of 0.24. Positive associations between g and the amount of gray and white matter have also been reported, although the associations regarding the latter seem to be slightly lower and less consistent (Narr et al., 2007; Posthuma et al., 2002). Finally, studies have also reported positive associations between general intelligence and the cortical thickness of specific brain areas (Narr et al., 2007; Shaw et al., 2006).

However, the association between g and brain characteristics is complicated by the fact that brain size is correlated with body mass. In the context of sex differences, males and females differ in average body length and mass, which transcends to sex differences in brain morphology (Ankney, 1992; Gur et al., 1999). Specifically, males have larger brains than females ($d \approx 0.7$), even after controlling for body size (Allen, Damasio, Grawboski, Buss, & Zhang, 2003; Ankney, 1992; Burgaleta et al., 2012; Rushton & Ankney, 1996). Combining this fact with the empirically determined correlation between brain size and g would predict that males also have higher g, but the literature has been inconsistent on this topic (e.g., see Colom, Juan-Espinosa, Abad, & García, L. F., 2000; Halpern, 2013; Halpern & LaMay, 2000). Inasmuch as this inconsistency has been addressed, it has been argued that differences in brain size across the sexes do not translate to g because of other anatomical sex differences, such as the density of neurons (Witelson, Glezer, & Kigar, 1995), cortical thickness (Sowell et al., 2007), the ratio of gray to white matter (Gur et al., 1999), or the functional organization of the brain (Sanders, Sjodin, & De Chastelaine, 2002). All these factors may lead to equal levels of g despite differences in brain volume (for a discussion see, Halpern et al., 2007).

Lynn (1994) noted that many studies showing null results were based on children and early adolescents, and pointed out that males and females differ in their growth curves and other aspects of maturation. If some aspect of brain development hinges on the level of hormones, and, in particular, sex differences in androgens, sex differences in intelli-

gence should be absent or otherwise slight until the age of around 15. From that age on, however, differences will start to emerge and eventually lead to a sex difference in *g* estimated to abound one third of a standard deviation, which translates to 3 to 5 IQ points (Lynn & Irwing, 2002, 2004). Consistent with this idea, several studies indicate that sex differences in *g* are related to sex differences in brain size in adults (Ankney, 1992; Lynn, 1994).

Sex differences in brain morphology are well established (Allen et al., 2003; Ruigrok et al., 2014) and neurological sex differences are receiving increased attention (Cahill, 2016). Whether those differences are also accompanied with differences in *g* remains unclear. First, as noted above, there seems to be no consensus yet regarding the basic question whether there are sex differences in *g* in the first place. Colom et al. (2000) reported only negligible sex differences in general intelligence in two samples totaling 10,474 adult applicants to a private university. Iliescu, Ilie, Ispas, Dobrean, and Clinciu (2016) tested several large representative samples in Romania and categorized the sample into various age groups. They reported that significant sex differences in general intelligence only occurred in a few of those age groups, whereas in the majority of the groups no significant differences were found. Halpern and LaMay (2000) reviewed the literature and concluded that even though there are sex differences in specific cognitive abilities, there are no differences in general intelligence.

Second, even when assuming that there may be sex differences in *g*, it remains a debate whether those are directly related to sex differences in brain characteristics. For example, Escorial et al. (2015) matched males and females on intelligence scores and found that despite equal levels of intelligence in those groups, males still had significantly larger brains than females, which led them to conclude that sex differences in brain size do not translate into average intelligence differences. Using the same line of reasoning, one of the more recent studies on brain morphology and *g* is of particular interest. Burgaleta et al. (2012) used structural MRI data of one hundred undergraduate students, and replicated the aforementioned sex difference in total brain volume, operationalized as the sum of the gray and white matter volume. After correcting for body size, the brain volume of males was approximately 0.75 of a standard deviation larger than that of females, consistent with differences in general brain size and gray and white matter. Burgaleta et al. (2012) also found that the differences in brain size were accompanied by sex differences in a limited set of specific cognitive abilities, mainly spatial ones. The sex difference in *g* was 0.12 in favor of males, which is small but not trivial, but it was not statistically significant. Altogether, Burgaleta et al. (2012) concluded that sex differences in brain morphology (i.e., size, proportion of gray and white matter) are not related to differences in *g*, but are instead associated with sex differences in specific cognitive abilities such as spatial ability, which, they argued, is facilitated by total brain volume in particular. Yet, Burgaleta et al. (2012) acknowledged several limitations that deserve attention. First, they used a sample of undergraduate psychology students only, which is not representative for the general population and which is likely to suffer from range restriction in *g*. Second, although their sample size (*N* = 100) was relatively large for a typical MRI study, it is quite small in absolute sense and may not have enough power to detect differences of the relevant size.

Thus, (1) Burgaleta et al. (2012) reported a small but non-significant sex difference in *g*, as well as a sex difference in brain size, and (2) Escorial et al. (2015) found that even in subsamples in which males and females have equal average intelligence scores, differences in brain size remain. Given these indeterminate results and conclusions, the question whether sex differences in *g* relate to sex differences in brain size remains open and motivates further evaluation with larger and more representative samples.

To this end, we analyzed the brain imaging findings and cognitive ability tests of the Human Connectome Project (HCP; Marcus et al., 2013; Van Essen, et al., 2013). The HCP is a large collaborative study designed to understand the neurobiology associated with an array of psychological variables, including cognitive ability. One of the advantages is that the most recent release of the HCP data comprises approximately 900 participants, which is a very large sample in comparison to other MRI studies. Although the HCP data still have a somewhat restricted age-range (22 to 37 years), the sample is demographically diverse and can be considered a more population-representative sample compared to that of Burgaleta et al. (2012). The HCP data set includes measures of intracranial volume and gray and white matter volume, which can be used to calculate total brain volume. Moreover, it provides measures of cortical thickness as well as cortical surface area. The HCP data also contains an array of ten cognitive tests, suitable for computing a *g* factor.

Beyond the relevance of testing brain and *g* sex differences in one of the largest MRI-datasets available, the present study contributes to the existing literature in the following ways. First, the detailed morphological data allow us to not only test for sex differences in brain characteristics and *g*, but also directly test whether brain characteristics statistically mediate any relation between sex and *g*. Second, we apply the Method of Correlated Vectors (MCV; Jensen, 1998). The MCV can be used to test to which extent the 'g dependence' of specific cognitive tasks relate to other variables. As such it allows to examine the extent to which any sex differences in the specific cognitive tasks occur as a function of *g* and brain size. Third, we apply a similar method as Escorial et al. (2015) in which a subsample of males and females who are matched on intelligence is tested on sex differences in brain size.

2. Method

2.1. Sample

The most recent release of the HCP contains 970 participants (HCP; 900 subject + MSM-All Reference Manual and Appendices, 2015) of which MRI and behavioral data is present from 896 participants, 393 males and 503 females. While the final sample is scheduled to include 1200 participants, data collection is ongoing and data are released in batches. Stringent selection procedures were followed to ensure that the participants were both physically and mentally healthy (Marcus et al., 2013; Van Essen, et al., 2013). The age of the participants ranged from 22 to 37 with a mean age of 28.82 years (*SD* = 3.66).

2.2. Brain imagine acquisition and processing and brain measures included

At Washington University, the brain magnetic resonance images (MRI) of the participants were obtained with a 3.0 Tesla (Siemens Skyra) scanner using a 32-channel head coil. The present study utilized the MRI imaging data that had been processed by the HCP team, using the following procedure. T1-Images were obtained with the T1-weighted 3D MPRAGE sequence using the following parameters: repetition time = 2400 ms; echo time = 2.14 ms; inversion time = 1000 ms; flip angle = 8°; field of view = 224 mm; matrix = 320 × 320; voxel size = 0.7 mm³ (Marcus et al., 2013; Van Essen et al., 2013). For the T2-weighted images, the 3D T2-Space sequence used the following parameters: repetition time = 3200 ms; echo time = 565 ms; the flip angle, field of view, matrix, and voxel size setting were identical to the ones used for the T1-weighted images.

For image processing, the HCP FreeSurfer processing pipeline was used (Glasser et al., 2013). Image processing was conducted using Talairach transformation, pial surface creation, skull registration, extraction of subcortical region volume, subcortical region segmentation, down sampling of T1-weighted images from 0.7 to 1 mm, and

intensity normalization (Glasser et al., 2013). The T2-weighted images were used to increase the accuracy of the pial surfaces and for removal of dural blood vessels.

Total brain size was calculated as the sum of the total amount of gray and white matter, identical to the procedure used by Burgaleta et al. (2012). Further, the gray/white matter ratio was calculated and direct measures of intracranial volume were also included in the analyses. In the HCP data thickness and surface area measures are available for 34 separate sections of the cortex for the right and left hemisphere. The distinction of the 34 areas is based on descriptions provided by Desikan et al. (2006) and more recently by Yeo, Ryman, Pommy, Thoma, and Jung (2016). Cortical thickness was operationalized as the average of the *Z*-transformed thickness measures of the 68 areas.

2.3. Cognitive measures

Participants of the Human Connectome Project sample conducted a wide range of cognitive validated tests derived from the NIH (National Institute of Health Services)-toolbox (http://www.nihtoolbox.org/; Weintraub et al., 2013) and from other sources. Some of the tests in the HCP were directly aimed at measuring fluid or crystallized intelligence, whereas others focused on more specific cognitive processes (e.g., executive functions, working memory, processing speed), which are nevertheless known to relate to general intelligence (Kane & Engle, 2002). A detailed description of each cognitive test and its scoring can be found at the HCP website (http://www.humanconnectome.org/documentation/Q1/behavioral-measures-details.html). In addition, a brief overview description of the tests used in the present study is provided below.

2.3.1. Penn progressive matrices (PPM)

The PPM (Bilker et al., 2012) is a short test of fluid intelligence, of similar construction as Raven's progressive matrices, one of the most well-known tasks to measure fluid intelligence. Participants are presented with arrangements of 2×2, 3×3, or 1×5 squares with patterns, except for one missing pattern. Participants must select from another set of five patterns the one that best fit the missing square in the arrangement. The PPM has 24 items and 3 additional determination items that increase in difficulty, and is discontinued if five incorrect responses are made in sequence.

2.3.2. Vocabulary size

Vocabulary size is an important aspect of crystallized intelligence. The present test is based on the Peabody Vocabulary test (Dunn & Dunn, 2007) and part of the NIH toolbox, and assesses word knowledge by presenting trials in which participants listen to an audio recording of a word and then have to select which of four pictures (presented on a computer screen) that best matches the word.

2.3.3. Oral reading recognition test

This (NIH Toolbox) test assesses vocabulary comprehension and reading decoding. The participants are asked to read and pronounce a set of letters and words as accurately as possible. The details of this task are described in Weintraub et al. (2013).

2.3.4. Working memory

Working memory is an executive function that in the literature has shown to overlap substantially with *g* (Kane & Engle, 2002; Salthouse, 2005). In the HCP, working memory is assessed with the List Sorting task from the NIH Toolbox, in which participants have to sequence various visually and orally presented stimuli in the categories Food and Animals. In one condition, participants have to order a series of objects

from the Food category in size from smallest to largest. In the other condition, they have to do the same, but stimuli are presented from both categories while participants have to order the stimuli within each category.

2.3.5. Episodic memory

Episodic memory was measured with the picture sequence memory task from the NIH Toolbox (Gur et al., 2010). In this task, series of visually presented objects and activities with increasing length are presented in a particular order. Participants are asked to recall the sequence of pictures. Sequence length varied from 6 to 18 pictures.

2.3.6. Spatial orientation

The short Penn line orientation test (Gur et al., 2010) was used to measure spatial orientation. The task consists of trials in which participants are presented with two lines in a different orientation. Participants have to move one of the lines so that it is parallel to the other.

2.3.7. Card sorting

The dimensional change card sorting (Zelazo, 2006) is an NIH Toolbox test that assesses executive functioning/cognitive flexibility. Participants are presented with two target cards that vary along the dimensions shape and color. From a set of test stimuli, participants have to select one that matches the relevant dimension of the current target card. During the task, the target dimension regularly changes, which requires cognitive flexibility to adjust one's card sorting strategy.

2.3.8. Verbal episodic memory

The word memory test (NIH Toolbox) assesses verbal episodic memory (see Weintraub et al., 2013 for details). Participants are first shown 20 words and are then asked which one they remember. Then a list of 40 words is presented, of which 20 are new and 20 the same as in the previous list. Participants then have to indicate whether they recognize the word in the 40 word list by choosing among the response options 'definitely yes', 'probably yes', 'probably no', and 'definitely no'.

2.3.9. Processing speed

Processing speed was measures with a task based on Salthouse Pattern Comparison Task (Salthouse, Babcock, & Shaw, 1991) Participants decide as quickly and accurately as possible whether an array of two side-by-side pictures are identical or not. The participant's raw score is the number of items correct in a 90-s period.

2.3.10. Flanker task

Adequate performance on the Flanker task requires both inhibition and sustained attention, two executive functions. The Flanker task from the NIH-toolbox is the Eriksen Flanker task as derived from the attentional network test (Rueda et al., 2004). Participants have to decide which of the two directions (left, right) an arrow presented in the middle of the screen point at. Each arrow is flanked by two other arrows on each side which are either in the same direction as the target arrow (congruent trials), or in the opposite direction (incongruent trials). The score on the version of this NIH Toolbox test is based on accuracy and reaction time.

2.4. Operationalization of g

Table 1 shows the zero-order intercorrelations between the cognitive tests, ranging from 0.05 to 0.68, with a mean of 0.243. These values are typical for the positive manifold of cognitive tests as has frequently been reported in the literature (e.g., Jensen, 1998).

First, in line with previous studies, we extracted the first unrotated

Table 1
Zero-order correlations between cognitive tasks.

	1	2	3	4	5	6	7	8	9
1. Progressive matrices	–								
2. Vocabulary	0.424	–							
3. Reading	0.432	0.675	–						
4. Working memory	0.345	0.329	0.338	–					
5. Episodic memory	0.281	0.169	0.162	0.332	–				
6. Spatial orientation	0.430	0.341	0.392	0.281	0.198	–			
7. Card sorting	0.223	0.182	0.248	0.164	0.208	0.250	–		
8. Verbal episodic mem	0.164	0.230	0.248	0.134	0.233	0.152	0.152	–	
9. Processing speed	0.134	0.144	0.140	0.179	0.183	0.129	0.411	0.088	–
10. Flanker task	0.141	0.211	0.186	0.145	0.144	0.196	0.516	0.054	0.405

Table 2
Factor loadings of the g factor on the cognitive tasks.

	Factor loadings overall g factor	Factor loadings restricted g factor
1. Progressive matrices	0.609	0.567
2. Vocabulary	0.664	0.790
3. Reading	0.694	0.807
4. Working memory	0.501	0.455
5. Episodic memory	0.394	
6. Spatial orientation	0.547	
7. Card sorting	0.477	
8. Verbal episodic memory	0.361	
9. Processing speed	0.357	
10. Flanker task	0.406	

Table 3
Mean values (standard deviations) and effect sizes (Cohen's d) for the brain characteristic variables by participant sex.

	Males	Females	d
Total brain size (mm^3)	1,211,846 (99,300)	1,061,487 (82,501)	0.57*
Intracranial volume (mm^3)	1,702,827 (148,397)	1,470,538 (149,672)	0.58**
Gray matter (GM) volume (mm^3)	731,507 (55,424)	646,298 (47,447)	0.57**
White matter (WM) volume (mm^3)	480,339 (50,454)	415,188 (40,698)	0.49**
GM/WM ratio	1.53 (0.10)	1.56 (0.10)	− 0.10*
Cortical thickness (Z score)	0.26 (1.10)	− 0.02 (0.91)	0.05
Cortical area (mm^2)	185,254 (15,470)	163,391 (13,220)	0.87**
g (Z score)	0.14 (0.89)	− 0.11 (0.88)	0.25**

* $p < 0.05$.
** $p < 0.01$.

factor (using principal axis factoring) from the entire set of cognitive measures. This factor had an Eigenvalue of 3.01 and explained a total of 33.40% of the variance (25.72% when considering the shared variance only). Table 2 shows the factor loadings of the general factor on each of the tests, which ranged from 0.32 to 0.69, with a mean of 0.49.

Second, we conducted the same procedure again but this time only included the limited set of cognitive tests that were developed to directly measure crystallized or fluid intelligence, namely the Penn Progressive matrices, the Vocabulary test, the reading test, and the working memory test. This first unrotated factor had an Eigenvalue of 2.29, explained 57.24% of the variance (45.07% shared variance), and showed factor loadings on the subtest ranging from 0.46 to 0.81 (see Table 2), with a mean loading of 0.66.

Notably, although the two operationalizations of g described above differed in their level of inclusiveness of cognitive tests, they correlated 0.90 with each other across participants, indicating that they were nearly identical. For a simpler presentation we report only the results of the g factor based on all cognitive tests, while asserting that the results with the restricted g factor were practically identical and all conclusions remained the same. The main results obtained with the restricted g factor are described in Appendix A.

3. Results

As mentioned in the Introduction, the literature reports substantial differences in brain morphology between males and females, which need to be considered when evaluating the relation between morphology and g. The means, standard deviations, and effect sizes of the brain measures for males and females are reported in Table 3. We first describe and test the sex differences and correlations between the intelligence and morphological measures. We then perform three different types of analyses to further examine the nature of these

relationships by means of (1) the method of correlated vectors, (2) statistical mediation analysis, and by (3) examining brain size differences between subsamples of participants matched on g.

First, analyses of variance (ANOVA) were conducted for each of the brain measures, partialling out body height (in cm). These analyses showed that, with the exception of cortical thickness, there was a significant sex difference for each measure. Males had significantly larger total brain volume ($F_{(1, 892)} = 157.51$, $p < 0.001$, $\eta_p^2 = 0.15$) and intracranial volume ($F_{(1, 892)} = 164.27.51$, $p < 0.001$, $\eta_p^2 = 0.16$), and had more gray and white matter ($F_{(1, 892)} = 157.83$, $p < 0.001$, $\eta_p^2 = 0.15$, and $F_{(1, 892)} = 115.70$, $p < 0.001$, $\eta_p^2 = 0.12$, respectively). Females scored significantly higher on the GM/WM ratio ($F_{(1, 892)} = 4.83.51$, $p < 0.05$, $\eta_p^2 = 0.01$). Males scored significantly higher on cortical surface area ($F_{(1, 892)} = 130.70$, $p < 0.001$, $\eta_p^2 = 0.13$), but, as mentioned above, there was no significant sex difference in cortical thickness ($F_{(1, 892)} = 1.02$, $p = 0.31$, $\eta_p^2 = 0.001$).

Importantly, the majority of the brain measures were significantly related to g, with the exception of GM/WM ratio for females and cortical thickness for both sexes. Table 4 shows the partial correlations, controlling for body height, between g and the seven brain volume measures for males, females, and the total sample (Appendix B contains the scatterplots of the correlations, with the regression lines for males and females).

The specific effects sizes regarding brain volume are consistent with

Table 4
Partial correlations (controlling for body height) of brain characteristics measures with the g factor.

	Overall sample	Males	Females
Total brain size	0.239**	0.245**	0.264**
Intracranial volume	0.257**	0.248**	0.293**
Gray matter (GM) volume	0.284**	0.305**	0.303**
White matter (WM) volume	0.116**	0.145**	0.180**
GM/WM ratio	0.082*	0.106*	0.066
Cortical thickness	0.012	0.024	−0.003
Cortical area	0.277**	0.289**	0.297**

* $p < 0.05$.
** $p < 0.01$.

those reported in recent meta-analyses (McDaniel, 2005; Pietschnig et al., 2015). For example, the overall correlation between brain size and intelligence was 0.24, which is identical to the recently meta-analytic established brain size-intelligence correlation reported by Pietschnig et al. (2015). Moreover, the findings in Table 4 shows that the associations between intelligence and brain characteristics are quite similar for males and females, except perhaps for the GM/WM ratio.

A one-way ANOVA showed that there was a significant difference in g ($F_{(1, 885)} = 17.12, p < 0.001, \eta_p^2 = 0.02$), with males scoring higher with an effect size of one-fourth of a standard deviation ($d = 0.25$). A subsequent MANOVA in which we included each of the separate cognitive tests was also significant ($F_{(10, 876)} = 11.22, p < 0.001, \eta_p^2 = 0.11$). Follow-up univariate tests showed that there were statistically significant sex differences on nine of the ten cognitive subtests, with the exception of processing speed (Table 5). Females scored higher on the episodic and verbal episodic memory tests, and males scored higher on the progressive matrices, vocabulary, reading, working memory, spatial orientation, and the Flanker task.

It can be assumed that the performance on the each of the cognitive tasks is based on a combination of the effects of g and any specific cognitive abilities, beyond g, such as spatial or verbal processing. Therefore, it is useful to test to which extent sex differences in cognitive abilities remain after removing the effects of g. The multivariate test was significant, ($F_{(9, 876)} = 10.83, p < 0.001, \eta_p^2 = 0.10$). The univariate statistics are displayed in Table 6 and reveal that, beyond g, males still scored significantly higher on spatial processing and the Flanker task. Females still scored significantly higher on the two episodic memory tasks. The other tasks did no longer show significant sex differences beyond those that can be attribute to g.

Table 5
Standardized scores for males and females for each cognitive tests, and the sex differences effect sizes, F-values, and p-levels.

	Means (SD)		d	$F_{(1, 884)}$	p
	Males	Females			
1. Progressive matrices	0.12 (1.00)	−0.09 (0.98)	0.21	9.09	0.003
2. Vocabulary	0.11 (1.01)	−0.09 (0.98)	0.20	8.23	0.004
3. Reading	0.13 (1.03)	−0.10 (0.97)	0.23	10.65	0.001
4. Working memory	0.10 (0.98)	−0.08 (1.02)	0.18	7.13	0.008
5. Episodic memory	−0.15 (0.99)	0.12 (0.99)	−0.27	16.28	< 0.001
6. Spatial orientation	0.24 (0.95)	−0.18 (1.00)	0.43	39.91	< 0.001
7. Card sorting	0.09 (1.02)	−0.07 (0.98)	0.16	5.76	0.017
8. Verbal episodic mem	−0.11 (1.02)	0.09 (0.97)	−0.20	8.86	0.003
9. Processing speed	0.06 (1.05)	−0.03 (0.95)	0.09	1.79	0.181
10. Flanker task	0.18 (1.05)	−0.13 (0.93)	0.26	21.77	< 0.001

Table 6
Standardized scores for males and females for each cognitive tests, and the sex differences effect sizes, F-values, and p-levels (controlling for the g factor).

	Means (and SDs)		d	$F_{(1, 884)}$	p
	Males	Females			
1. Progressive matrices	0.01 (1.01)	−0.01 (0.99)	0.02	9.09	0.770
2. Vocabulary	−0.01 (0.99)	0.01 (1.00)	0.02	8.23	0.794
3. Reading	0.01 (1.03)	−0.0 (0.98)	0.01	10.65	.0.924
4. Working memory	0.02 (0.98)	−0.01 (1.02)	0.03	7.13	0.678
5. Episodic memory	−0.24 (1.00)	0.19 (0.95)	−0.43	16.28	< 0.001
6. Spatial orientation	0.18 (0.99)	−0.14 (0.98)	0.32	39.91	< 0.001
7. Card sorting	0.01 (1.02)	−0.01 (0.98)	0.02	5.76	0.809
8. Verbal episodic mem	−0.18 (1.03)	0.14 (0.95)	−0.32	8.86	< 0.001
9. Processing speed	−0.01 (1.07)	0.01 (0.94)	0.02	1.79	0.743
10. Flanker task	0.12 (1.05)	−0.10 (0.95)	0.22	21.77	0.002

3.1. Method of correlated vectors

To gain further insight into the nature of the cognitive differences, we applied the method of correlated vectors (MCV). Some scholars have pointed to limitations of this method (Ashton & Lee, 2005) such as sensitivity for invariance and the influences of specific abilities (beyond g). However, as the factor loadings of the specific cognitive tasks as well as the brain-g correlations were rather similar in males and females, invariance is unlikely to pose a problem in the present sample. The potential diminishing effects of specific cognitive abilities on the vector correlation can also be expected to be relatively low, given that g played a strong role in the cognitive tasks (see Table 6). Further, to date the MCV remains to be a widely used method in studies aiming to identify which variables are associated with g (see for example Te Nijenhuis, Jongeneel-Grimen, & Armstrong, 2015; Woodley of Menie, Te Nijenhuis, Frenandes, & Metzen, 2016).

The MCV was applied in three ways. First, we correlated (using Spearman's r) the g loading of the ten cognitive tasks (see Table 2) with their sex-difference effect sizes. This correlation was $r_{(10)} = 0.67$ ($p = 0.03$) showing that the sex differences were larger for those tasks that showed the strongest g loadings. Second, we correlated the tasks' g loadings with their correlation with brain size. This correlation was $r_{(10)} = 0.84$ ($p = 0.002$), showing that the correlation between cognitive ability and brain size increased with the g loading of the cognitive task. Finally, we correlated the sex-difference effect sizes of the ten tasks with their correlation with brain size. In this case the correlation was $r_{(10)} = 0.85$ ($p = 0.002$), showing that the sex differences were larger as a function of their relation to brain size.

3.2. Mediation analysis of brain characteristics on g differences

Since the data showed significant sex differences in both brain characteristics and g, a critical question was whether morphological differences mediate the group difference in g. Although the MCV already indicated that this was the case, we included mediation analyses to examine for consistency across analysis methods. The likelihood that one variable mediates another in a cross-sectional design can be tested. Here, we used the SPSS PROCESS module (v 2.16) of Hayes (2013) with the bootstrap method and 5000 iterations. First, in order to use total brain volume as a mediation variable, we started with regressing body height on brain size and saving the residuals. These residuals reflect the sex difference in total brain volume that cannot be attributed to sex differences in physical height. Subsequently, we tested mediation of residualized brain size on the
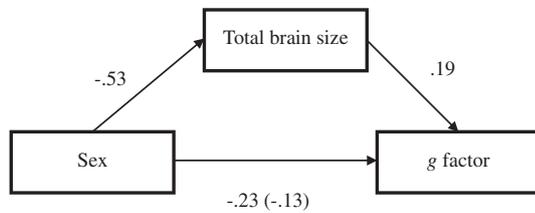
Fig. 1. Mediation test of (residual) total brain size on the relation between sex and *g*.

relation between sex and *g*. The indirect (mediation) effect was significant ($B = -0.103$, $SE = 0.021$, $Z = -4.86$, $p < 0.001$, Confidence Interval Low (CIL) $= -0.144$, Confidence Interval High (CIH) $= -0.061$). The ratio of the indirect effect to the total effect of sex on *g* suggests that approximately 43% (($0.53 * 0.19$) / 0.23) of the relation between sex and *g* is mediated by differences in total brain volume, as depicted in Fig. 1.

Similar significant mediation effects were found for cortex surface area ($B = -0.114$, $SE = 0.021$, $Z = -5.37$, $p < 0.001$, $CIL = -0.158$, $CIH = -0.076$; the indirect effect accounted for 42% of the total effect of the sex-*g* difference), intracranial capacity ($B = -0.107$, $SE = 0.021$, $Z = -5.10$, $p < 0.001$, $CIL = -0.154$, $CIH = -0.069$; the indirect effect accounted for 47% of the total effect of the sex-*g* difference), gray matter ($B = -0.113$, $SE = 0.021$, $Z = -5.29$, $p < 0.001$, $CIL = -0.157$, $CIH = -0.075$; the indirect effect accounted for 49% of the total effect of the sex-*g* difference), and white matter ($B = -0.061$, $SE = 0.017$, $Z = -3.48$, $p < 0.001$, $CIL = -0.110$, $CIH = -0.032$; the indirect effect accounted for 47% of the total effect of the sex-*g* difference). The indirect effect of the GM/WM ratio did not reach the 0.05 significance level ($B = 0.015$, $SE = 0.008$, $Z = 1.80$, $p = 0.07$, $CIL = 0.002$, $CIH = 0.037$) indicating that the GM/WM ratio does not statistically mediate the sex differences in *g*.

3.3. Sex differences in brain size after matching on g

A final test of the relation between brain size and sex differences in cognitive ability followed the procedure as reported by Escorial et al. (2015). To this end, from the present participants, we selected a male and a female subsample matched on their *g* score. The subsample consisted of 680 participants (340 males, 340 females) in which, due to the matching procedure, males and females showed nearly identical *g* scores ($F_{(1, 678)} = 0.001$, $p = 0.97$, $\eta_p^2 = 0$). Examining the total brain volume in these two matched groups showed that the sex differences in brain size (controlling for height) remained highly significant ($F_{(1, 676)} = 116.81$, $p < 0.001$, $\eta_p^2 = 0.147$). Moreover, the effect size was nearly identical compared to the whole sample (which was $\eta_p^2 = 0.150$, see above). These findings are in line with those reported by Escorial et al. (2015) who drew the conclusion that because there remained a sex difference in brain size even when *g* was identical, brain size is unrelated to cognitive ability. The basis for this conclusion was that females ostensibly achieve a similar level of intelligence with a smaller average brain size.

For several reasons, however, we consider it unlikely that those matching results support the conclusion that brain size is not related to the group difference in *g*. First, given that there is a well-documented correlation between *g* and brain size both within and across the sexes it is likely that brain size differences affect general intelligence and does so within and across sexes. Second, the conclusion may not be statistically valid, because there are various physical, behavioral, and cognitive aspects on which males and females may differ and that can contribute to sex differences in brain size (Cahill, 2016). Therefore,

matching males and females on one specific variable does not necessarily imply that brain size differences will greatly diminish, particularly when the correlation between that variable with brain size is less than unity.

In accordance with these two latter contemplations, we decided to examine their implications empirically for males and females separately. Specifically, we first divided the group of male participants into two groups high and low in total brain volume (controlling for height), based on a median split procedure. Because brain size and *g* correlate, we expected the high brain volume group to have a higher *g* score than the low volume group. Second, identically to the procedure applied to the sex difference, a subsample was extracted in which the *g* scores of participants in the high brain volume group were matched with participants in the low brain volume group, which is the analogue of matching males and females on *g*. These procedures were applied to the female participants as well, and the findings are reported for males and females separately below.

For the males, obviously, due to the median-split procedure, the group with brain sizes below the median scored significantly lower on brain size (controlling for height) than the group above the median ($F_{(1, 390)} = 655.20$, $p < 0.001$, $\eta_p^2 = 0.627$). Importantly, however, the two groups also significantly differ in *g* score ($F_{(1, 389)} = 17.27$, $p < 0.001$, $\eta_p^2 = 0.043$). As the only criterion on which the groups were selected was brain volume, it seems reasonable to conclude that the differences in the two groups of males were related to their differences in brain size. Subsequently, we tested whether the males in the two groups still differed in brain volume in the subsample of *g*-score matched participants. First, the success of the matching procedure became apparent in the fact that the two subgroups of males no longer significantly differed in their *g* score ($F_{(1, 288)} = 0$, $p = 0.99$, $\eta_p^2 = 0$). Nevertheless, the groups remained significantly different in mean brain volume, and the effect size was nearly identical to the full sample of males ($F_{(1, 287)} = 496.54$, $p < 0.001$, $\eta_p^2 = 0.634$).

The above described pattern of findings was similar in the females. Females in the below-median group had less brain volume as well as lower *g* scores than those in the above-median group ($F_{(1, 499)} = 847.40$, $p < 0.001$, $\eta_p^2 = 0.629$ and $F_{(1, 494)} = 30.30$, $p < 0.001$, $\eta_p^2 = 0.058$, respectively). Likewise, the females in the two groups matched on *g* did no longer differ in *g* score ($F_{(1, 360)} = 0$, $p < 1$, $\eta_p^2 = 0$). However, in this subsample, the differences in brain size remained significant with almost identical effect sizes compared to the full sample of females ($F_{(1, 358)} = 571.85$, $p < 0.001$, $\eta_p^2 = 0.615$).

These additional analyses of males and females in groups based on brain volume provide valuable information because they reveal that the paradox of brain size and intelligence as reported by Escorial et al. (2015) does not only hold when comparing males and females, but is similar in within-sex comparisons in which participants have been allocated based on their brain volume.

4. Discussion

The present study examined the associations between brain characteristics and *g*, utilizing the dataset from the HCP that includes MRI data of almost 900 participants. First, measures of total brain volume, level of gray and white matter, intracranial volume, and cortical surface area all were significantly and positively associated with *g*, in both males and females. The effects sizes were quite similar to those reported in previous studies, such as the meta-analytical estimate of 0.24 (Pietschnig et al., 2015). The brain size-*g* relationship was also confirmed by the MCV, showing that those cognitive subtasks with the highest *g* loadings showed the strongest correlations with brain size (Woodley of Menie et al., 2016). Further, applying a median-split procedure on brain size for males and females separately clearly

supported the notion that, for both sexes, bigger brains are accompanied with higher average general intelligence scores.

The GM/WM ratio and cortical thickness did not correlate significantly with *g* scores. Previous studies have already noted that the GM/WM ratio show inconsistent results regarding its relation with *g* (e.g., Burgaleta et al., 2012). A possible reason for the null finding regarding the cortical thickness-*g* association is that the thickness of specific brain areas, rather than an average, relate to general intelligence (Haier, Jung, Yeo, Head, & Alkire, 2005; Narr et al., 2007). This is further complicated by the fact that the cortical thickness in its relation to cognitive ability seems to depend on sex, body size, and age (Plessen, Hugdahl, Bansal, Hao, & Peterson, 2014).

Second, the present study confirmed that males have larger total brain volume, intracranial volume, and amount of gray and white matter, both before and after controlling for body size (Lüders, Steinmetz, & Jäncke, 2002; Lynn, 1994; Nyborg, 2005; Rushton & Ankney, 2009). The effect sizes were in accordance with those reported in the recent MRI study of Burgaleta et al. (2012), and have in fact been mentioned in many older studies too (Ankney, 1992; Rushton & Ankney, 1996).

Females were found to have a higher gray to white matter ratio. Previous findings on the GM/WM ratio are inconsistent with some studies reporting significant sex differences favoring females (e.g., Gur et al., 1999), whereas others did not find such a difference (e.g., Burgaleta et al., 2012). While it has been suggested that a higher GM/WM ratio may reflect a higher level of processing efficiency (Gur et al., 1999), there is currently no clear functional interpretation of this ratio (Burgaleta et al., 2012). Moreover, similar to the findings in the present study, the GM/WM ratio often shows non-significant or weak relations with cognitive ability scores. Also, no significant sex differences in overall cortical thickness were found. This finding is consistent with several previous studies (Crespo-Facorro et al., 2011; Escorial et al., 2015; White, Andreasen, Nopoulos, & Magnotta, 2003), but inconsistent with others (Im, Lee, Lyttelton, Kim, & Evans, 2008; Sowell et al., 2007).

Third, there was a significant sex difference in *g* in this sample, with an effect size of one quarter of a standard deviation. This corresponds to approximately 3.75 IQ points, which is a similar sex difference in general intelligence as reported in previous large population studies (Lynn & Irwing, 2004) and meta-analyses (e.g., Irwing & Lynn, 2005; Madison, 2016). Analyses of the cognitive subsets used to extract *g* showed that the sex differences were not related to extremely high scores of males on a limited number of particular tasks, but tends rather to reflect a more general pattern. Moreover, controlling for the effect of *g*, showed that the majority of sex differences in favor of males were no longer significant, which further supports the notion that *g* played an important role in causing the initial differences. The tasks that remained to display significant sex differences beyond *g* were the spatial orientation task, which also strongly taps spatial ability, and the Flanker tasks that measures sustained attention and inhibition (Rueda et al., 2004).

Females scored consistently higher on the two episodic memory tasks (See also Bleecker, Bolla-Wilson, Agnew, & Meyers, 1988; Herlitz, Nilsson, & Bäckman, 1997; Herlitz & Rehnman, 2008). The *g* loadings of the two episodic memory tasks were modest, however, and probably therefore did not balance the sex differences on the latent factor.

The MCV results were in line with the notion of sex differences in *g*, because they showed that the sex differences were largest on those cognitive subtasks task that also had the highest *g* loadings. It is relevant to note that the present MCV results differ from those reported by Colom, García, Juan-Espinosa, and Abad (2002), who, based on a rather low vector correlation between sex differences and *g*, argued that sex difference were unrelated to general intelligence. A conclusive

answer regarding the different findings in the present study and those of Colom cannot be provided. Nevertheless, Ashton and Lee (2005) extensively discuss the MCV findings by Colom et al. (2002) and concluded that the effects of specific cognitive abilities (beyond *g*) may have diminished the vector correlation. They also concluded that even in the Colom et al. (2002) study is it likely that the sex differences did relate to *g*, which would be in line with the present findings.

Fourth, and the central point of the present study, is that the various statistical methods applied seem to suggest that sex differences in brain characteristics indeed mediate sex differences in *g*. Direct support for this notion came from the mediation analyses, indicating that brain volume measures could account for roughly half of the sex differences in *g*. It should be noted that the proportion of variance accounted for by mediation analysis constitutes a lower bound, because all unsystematic variability related to random error and poor reliability and validity is aggregated in the denominator. Indirect support for the involvement of brain size in sex differences in *g* came from the MCV, showing that sex differences were largest on those specific cognitive tasks that showed the strongest associations with brain size as well as the highest *g* loadings.

In this context, one finding needing more explanation is that a subset of males and females, matched on *g*, still exhibited a significantly different brain size. A similar phenomenon, for cortical area, was reported by Escorial et al. (2015) who concluded that "…observed sex differences in the analyzed neocortical measures are unrelated with the cognitive performance of men and women" (p.360) and that sex differences in brain size do not translate into sex differences in *g*. As mentioned in the Results section, however, in our view the fact that individuals with equal general intelligence can still differ in brain size, does not necessarily validate the conclusion that brain size is therefore unrelated to group differences in *g*. A large proportion of the brain is involved in producing other outcomes than general intelligence. Thus, even males and females, matched on general intelligence, may still differ on a range of other physical, behavioral and cognitive aspects that cause their brain size to differ. That this phenomenon is not exclusive for sex differences was demonstrated by the within-sex median split procedure applied. Specifically, even within each sex, participants categorized according to their brain size differ in their mean *g* scores. This clearly supports the notion that difference in brain size are responsible for the differences in *g* among the subgroups. Yet, when matched on *g* score, the differences in brain size remained largely the same.

The total pattern of outcomes of the AN(C)OVAs, MCV, the mediation analyses, and the median-split procedure warrant the conclusion that sexual dimorphic brain characteristics may not exclusively contribute to sex differences in specific abilities (e.g., episodic memory, spatial ability), as have previously been suggested (Halpern, 2013; Witelson et al., 1995), but contribute directly to sex differences in *g*. These findings and conclusions differ from those reported in Burgaleta et al. (2012), and it is useful to pay attention to the possible reasons for this discrepancy. One likely explanation is the restriction of range in the sample of Burgaleta et al.'s study, which consisted of 100 undergraduate students. The fact that all these students had chosen to study psychology implies that they have more similar interests and levels of specific abilities than males and females in the general population. In the sample of Burgaleta et al. the sex differences in *g* was in the expected direction, but was smaller (*d* = 0.12, corresponding to 1.8 IQ points) than has been reported in other population studies (Irwing & Lynn, 2005; Madison, 2016). Given that range restriction in intelligence was probably substantial among the psychology students, it can be expected that any correction for range restriction (e.g., Schmidt & Hunter, 2014) would have enlarged the effect to a size that may have been closer to the one found in the present HCP sample.

4.1. Limitations and future research

In the present study, several limitations need to be taken into account when interpreting the results. One is that the HCP data is cross-sectional and therefore the mediation and MCV analyses cannot directly infer causal relations between brain size and *g*, but rather indicate that there is overlap in the variance.

A second limitation was that the sample still had a certain level of range restriction in age (22–37 years) and mental and physical health levels (all participants were healthy). On the other hand, the effect sizes of the brain-*g* correlations were nearly identical to population-based meta-analytic values (McDaniel, 2005; Pietschnig et al., 2015). This suggests that the sample is reasonably representative for the purposes of the present study, in terms of the relation between morphology and cognitive ability.

A last limitation is that, although the focus of the present study was on general brain characteristics (e.g., size, area, thickness), there are more specific brain areas or structures that have shown particular relevance with regard to general intelligence, such as the prefrontal cortex and the hippocampus (Duncan et al., 2000). Future research on the potential role of such specialized areas in the sex difference in *g* would be useful. In the same vein, future research would benefit from not only considering structural MRI, but also include functional MRI (fMRI), which can reveal whether males and females differ in the brain areas and level of activation when engaging in tasks that tax general intelligence.

4.2. Concluding remarks

It is an interesting observation that in the nineteenth century the consensus was that sex differences in brain size exists, leading to a slightly higher average of males in general intelligence (e.g., Darwin, 1871). However, improved psychometric and brain imaging techniques have led to a new wave of studies and have reactivated the debate on this topic. Regarding this, the present study may contribute to this field by applying a combination of newer and more traditional methods. Overall, we agree with the conclusion of Burgaleta et al. (2012) and Escorial et al. (2015) that within subgroups or at the individual level, larger male brains do not necessarily have to be accompanied with higher general intelligence. Nevertheless, the present study also clearly indicates that, at the group level, there is a sex difference in *g* and that differences in brain size likely play a relevant role in this. Given that those conclusions were based on the results of one of the larger MRI studies available, it can be expected that the effect sizes provide reliable estimates of the relations and can be regarded as benchmarks in the literature in this area.

Appendix A: Results with the restricted g factor (based on four cognitive tasks)

In the manuscript, we referred to an alternative *g* factor based on a limited set of four cognitive tasks that were more directly designed to test aspects of intelligence. This $g_{(restricted)}$ was strongly related to the *g* factor that was based on the full set of (ten) cognitive tasks ($r = 0.90$). Moreover, in general, the findings and conclusions based on the $g_{(restricted)}$ are similar to those with the full *g* factor. Several of the basic analyses involving the $g_{(restricted)}$ are described below.

First, as with the full *g* factor, there was a significant sex difference on the $g_{(restricted)}$, ($F_{(1, 888)} = 10.92, p < 0.001, \eta_p^2 = 0.02$). Identical to the full *g* factor described in the paper, this amounts to an effect size of $d = 0.25$.

Table A1 describes the correlations between the $g_{(restricted)}$ with the various brain measures used in the study. As becomes apparent, the correlations are very similar to those obtained with the full *g* factor. For example, the overall effect sizes between total brain size and *g* were 0.24, 24, and 0.26 for the total sample, males, and females, respectively. The GM/WM ratio did no longer reach significance at the $p < 0.05$ level in the total sample and for the males. However, in absolute terms, even those correlations also resembled each other strongly (e.g., 0.063 vs. 0.082 and 0.066 vs. 0.106, for the total sample and males, respectively).

Table A1
Correlations between the $g_{(restricted)}$ and brain characteristics.

	Overall sample	Males	Females
Total brain size	0.239**	0.241**	0.264**
Intracranial volume	0.241**	0.235**	0.271**
Gray matter (GM) volume	0.277**	0.286**	0.304**
White matter (WM) volume	0.165**	0.158**	0.180**
GM/WM ratio	0.063*	0.066	0.064
Cortical thickness	0.010	0.010	0.006
Cortical area	0.275**	0.279**	0.300**

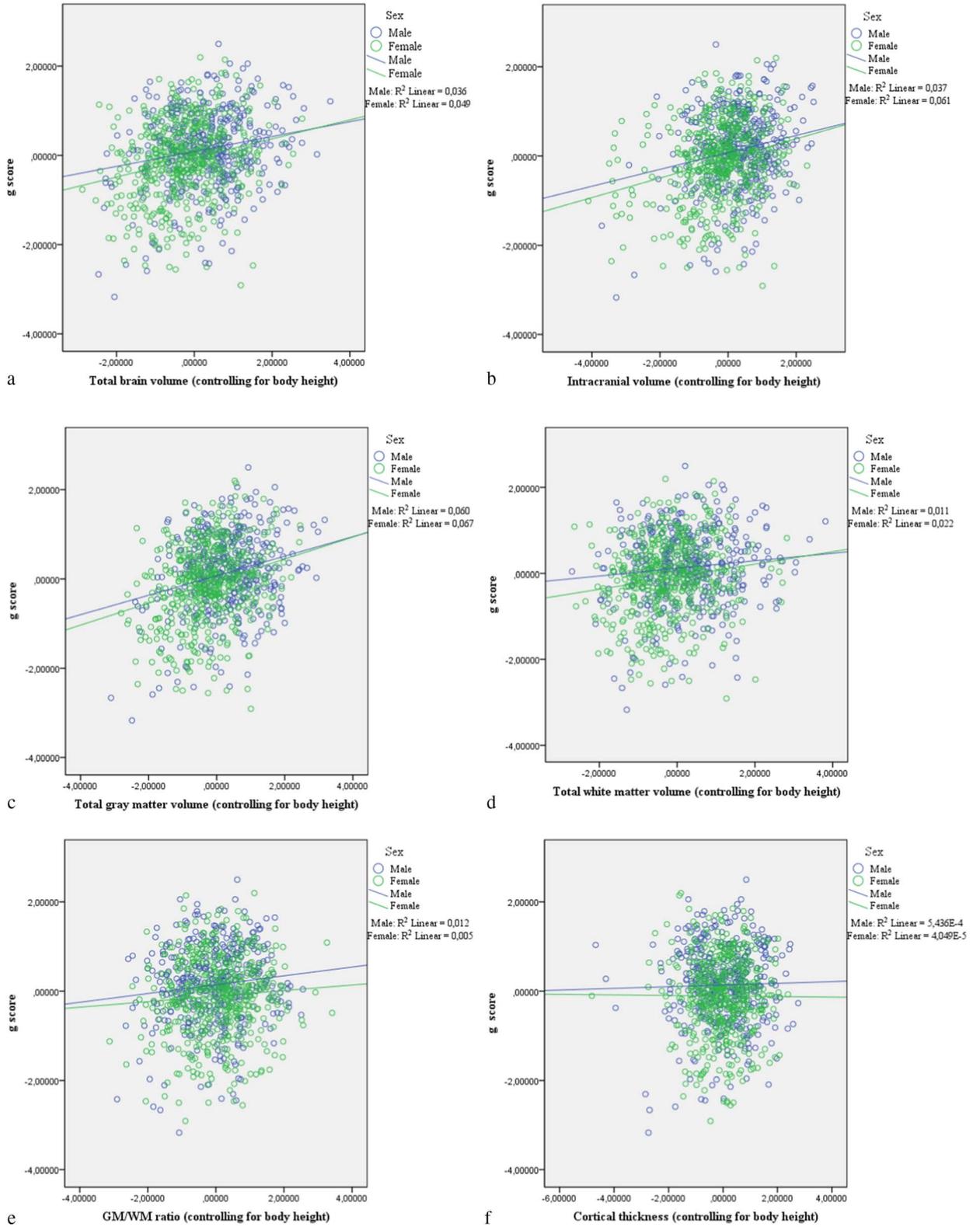* $p < 0.05$.
** $p < 0.01$.

Appendix B
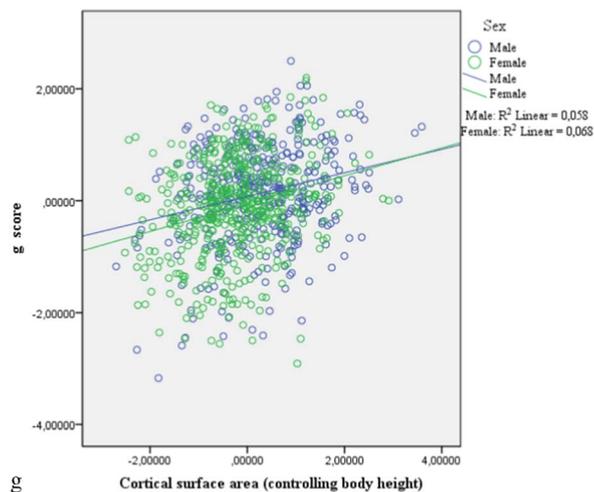

Fig. 1. Scatterplots Brain characteristics and *g*

Fig. 1. (Continued)

References

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33*, 431–444.

Allen, J. S., Damasio, H., Grawboski, T. J., Buss, J., & Zhang, W. (2003). Sexual dimorphism and asymmetries in the gray-white composition of the human cerebrum. *NeuroImage, 18*, 880–894.

Ankney, C. D. (1992). Sex differences in relative brain size: The mismeasure of woman too? *Intelligence, 16*, 329–336.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). *Development of the abbreviated nine-item forms of the Raven's Standard Progressive Matrices test.* (Assessment, 1073191112446655).

Bleecker, M. L., Bolla-Wilson, K., Agnew, J., & Meyers, D. A. (1988). Age-related sex differences in verbal memory. *Journal of Clinical Psychology, 44*, 403–411.

Burgaleta, M., Head, K., Álvarez-Linera, J., Martínez, K., Escorial, S., Haier, R., & Colom, R. (2012). Sex differences in brain volume are related to specific skills, not to general intelligence. *Intelligence, 40*, 60–68.

Cahill, L. (2016). An issue whose time has come. *Journal of Neuroscience Research, 95*, 12–13.

Colom, R., Juan-Espinosa, M., Abad, F., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence, 28*, 57–68.

Colom, R., García, L. F., Juan-Espinosa, M., & Abad, F. J. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *The Spanish Journal of Psychology, 5*, 29–35.

Colom, R., Haier, R. J., Head, K., Álvarez-Linera, J., Quiroga, M.Á., Shih, P. C., & Jung, R. E. (2009). Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence, 37*, 124–135.

Crespo-Facorro, B., Roiz-Santiáñez, R., Pérez-Iglesias, R., Mata, I., Rodríguez-Sánchez, J. M., Tordesillas-Gutiérrez, D., ... Magnotta, V. (2011). Sex-specific variation of MRI-based cortical morphometry in adult healthy volunteers: The effect on cognitive functioning. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 35*, 616–623.

Darwin, C. (1871). *The descent of man, and selection in relation to sex.* London: John Murray.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*, 968–980.

Duncan, J., Seitz, R. J., Kolodny, J., et al. (2000). A neural basis for general intelligence. *Science, 289*, 457–460.

Dunn, L., & Dunn, D. (2007). *PPVT-4: Peabody picture vocabulary test manual*(4th ed.). Minneapolis, MN: Pearson Assessments.

Escorial, S., Román, F. J., Martínez, K., Burgaleta, M., Karama, S., & Colom, R. (2015). Sex differences in neocortical structure and cognitive performance: A surface-based morpometry study. *NeuroImage, 104*, 355–365.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... Van Essen, D. C. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage, 80*, 105–124.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.

Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: Correlations with cognitive performance. *The Journal of Neuroscience, 19*, 4065–4072.

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., ... Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct

validation. *Journal of Neuroscience Methods, 187*, 254–262.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage, 25*, 320–327.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229–246.

Halpern, D. F., Benbow, C. P., Geary, C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1–51.

Halpern, D. F. (2013). *Sex differences in cognitive abilities.* Psychology press.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Press.

Herlitz, A., Nilsson, L. G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 801–811.

Herlitz, A., & Rehnman, J. (2008). Sex differences in episodic memory. *Current Directions in Psychological Science, 17*, 52–56.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage publications.

Iliescu, D., Ilie, A., Ispas, D., Dobrean, A., & Clinciu, A. I. (2016). Sex differences in intelligence: A multi-measure approach using nationally representative samples from Romania. *Intelligence, 58*, 54–61.

Im, K., Lee, J. M., Lyttelton, O., Kim, S. H., Evans, A. C., & Kim., S. I. (2008). Brain size and cortical structure in the adult human brain. Cerebral Cortex, 18, 2181–2191.

Irwing, P., & Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta-analysis. *British Journal of Psychology, 96*, 505–524.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport CT: Praeger.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences, 17*, 257–271.

Lynn, R., & Irwing, P. (2002). Sex differences in general knowledge, semantic memory and reasoning ability. *British Journal of Psychology, 93*, 545–556.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481–498.

Lüders, E., Steinmetz, H., & Jäncke, L. (2002). Brain size and grey matter volume in the healthy human brain. *Neuroreport, 13*, 2371–2374.

Madison, G. (2016). Sex differences in adult intelligence in Sweden. *Mankind Quarterly, 57*, 24–33.

Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., ... Hodge, M. (2013). Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage, 80*, 202–219.

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.

Narr, K. L., Woods, R. P., Thompson, P. M., Szeszko, P., Robinson, D., Dimtcheva, T., ... Bilder, R. M. (2007). Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cerebral Cortex, 17*, 2163–2171.

Nyborg, H. (2005). Sex-related differences in general intelligence g, brain size, and social status. *Personality and Individual Differences, 39*, 497–509.

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 & Biobehavioral Reviews, 57*, 411–432.

Plessen, K. J., Hugdahl, K., Bansal, R., Hao, X., & Peterson, B. S. (2014). Sex, age, and cognitive correlates of asymmetries in thickness of the cortical mantle across the life

span. *The Journal of Neuroscience, 34*, 6294–6302.

Posthuma, D., De Geus, E. J., Baaré, W. F., Pol, H. E. H., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience, 5*, 83–84.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42*, 1029–1040.

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews, 39*, 34–50.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychon. Bull. Rev. 3*(1), 21–36.

Rushton, J. P., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience, 119*, 692–732.

Sanders, G., Sjodin, M., & De Chastelaine, B. (2002). On the elusive nature of sex differences in cognition: Hormonal influences contributing to within-sex variation. *Archives of Sexual Behaviour, 31*, 145–152.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6*, 118–127.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature, 440*, 676–679.

Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., ... Toga, A. W. (2007). Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cerebral Cortex, 17*, 1550–1560.

Te Nijenhuis, J., Jongeneel-Grimen, B., & Armstrong, E. L. (2015). Are adoption gains on the g factor? A meta-analysis. *Personality and Individual Differences, 73*, 56–60.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium (2013). The WU-Minn human connectome project: An overview. *NeuroImage, 80*, 62–79.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., & Fox, N. A. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*(11 Suppl. 3), S54–S64.

White, T., Andreasen, N. C., Nopoulos, P., & Magnotta, V. (2003). Gyrification abnormalities in childhood-and adolescent-onset schizophrenia. *Biological Psychiatry, 54*, 418–426.

Wickett, J. C., Vernon, P. A., & Lee, D. H. (1994). In vivo brain size, head perimeter, and intelligence in a sample of healthy adult females. *Personality and Individual Differences, 16*, 831–838.

Witelson, S. F., Glezer, I. I., & Kigar, D. L. (1995). Women have greater density of neurons in posterior temporal cortex. *The Journal of Neuroscience, 15*, 3418–3428.

Woodley of Menie, M. A., Te Nijenhuis, J., Frenandes, H. B. F., & Metzen, D. (2016). Small to medium magnitude Jensen effects on brain volume: A meta-analytic test of the processing volume theory of general intelligence. *Learning and Individual Differences, 51*, 215–219.

Yeo, R. A., Ryman, S. G., Pommy, J., Thoma, R. J., & Jung, R. E. (2016). General cognitive ability and fluctuating asymmetry of brain surface area. *Intelligence, 56*, 93–98.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols-Electronic Edition, 1*, 297.