Prefrontal Cortex and Executive Functions in Healthy Adults: A Meta-Analysis of Structural Neuroimaging Studies

Peng Yuan[a] and **Naftali Raz**[a,*]

[a]Institute of Gerontology and Department of Psychology, Wayne State University, 87 East Ferry St., 226 Knapp Bldg., Detroit, MI 48202

Abstract

Lesion studies link the prefrontal cortex (PFC) to executive functions. However, the evidence from *in vivo* investigations in healthy people is mixed, and there are no quantitative estimates of the association strength. To examine the relationship between PFC volume and cortical thickness with executive cognition in healthy adults, we conducted a meta-analysis of studies that assessed executive functions and PFC volume (31 samples,) and PFC thickness (10 samples) *in vivo*, N=3272 participants. We found that larger PFC volume and greater PFC thickness were associated with better executive performance. Stronger associations between executive functions and PFC volume were linked to greater variance in the sample age but was unrelated to the mean age of a sample. Strength of association between cognitive and neuroanatomical indices depended on the executive task used in the study. PFC volume correlated stronger with Wisconsin Card Sorting Test than with digit backwards span, Trail Making Test and verbal fluency. Significant effect size was observed in lateral and medial but not orbital PFC. The results support the "*bigger is better*" hypothesis of brain-behavior relation in healthy adults and suggest different neural correlates across the neuropsychological tests used to assess executive functions.

Keywords

brain; MRI; morphometry; cognition; aging; volume; cortical thickness

Introduction

Executive control of cognition is a popular area of research in psychology and cognitive neuroscience, and typing "executive functions" into a PubMed search engine yields more than 4100 articles spanning almost five decades of scholarship (http://www.ncbi.nlm.nih.gov/pubmed/?term="executive+functions"; accessed on January 26, 2014). In spite of a prolonged effort, the understanding of this complex cognitive domain and its neural foundations continues to evolve. Much has been written about conceptual and

*Corresponding author, nraz@wayne.edu. Tel.: +1-313-664-2617.

measurement problems with the concept of executive functions including limited construct validity, poor-to-moderate reliability, and operational heterogeneity of the tasks used for their assessment (Miyake et al, 2000; Stuss and Alexander, 2000, see Alvarez and Emory, 2006 for a comprehensive review). Thus, integrating evidence over multiple samples may bring insights that elude the investigators in single studies.

Executive function as cognitive construct is not a matter of consensus, and the extant definitions reflect the diversity of the measurements that purport to quantify it as well as the priorities and specific interests of the research discipline from which they originate. Neuropsychologists base their definitions on observations of patient responses to brain insults and view executive functions as "*capacities that enable a person to engage successfully in independent, purposive, self-serving behavior*" (Lezak, 1995) and "*a variety of loosely related higher-order cognitive processes including initiation, planning, hypothesis generation, cognitive flexibility, decision making, regulation, judgment, feedback utilization, and self-perception that are necessary for effective and contextually appropriate behavior*" (Spreen and Strauss, 1998). Others add "*… judgment, working memory and shifting set…*" *that " enable one to … orient towards the future…"* (Cheung et al., 2004). A psychometric view maintains that "executive functions" is "*an umbrella term comprising a wide range of cognitive processes and behavioral competencies which include verbal reasoning, problem-solving, planning, sequencing, the ability to sustain attention, resistance to interference, utilization of feedback, multitasking, cognitive flexibility, and the ability to deal with novelty*" (Chan et al., 2008). Experimental psychologists have been following a definition of executive functions that stresses the central executive, a supervisor of specific cognitive operations (Baddeley, 1996). All agree, however, that executive functions represent multifarious cognitive phenomena. As such, they are unlikely to be "localized" to a specific area of the brain, although some brain circuits can still show a non-random association with executive cognition. Therefore, executive performance may be related to functional and structural characteristics of specific brain regions as well as the connections among them.

Traditionally, success on executive (or "higher cognitive") tasks has been attributed to activity and structural integrity of the frontal lobes (Luria, 1966; Teuber, 1972). So strong is the perceived connection between the function and the structure that in discussing tests, the terms "executive" and "frontal" are frequently used interchangeably. Such use is still prevalent, in spite of well-reasoned critique of this practice (Alvarez and Emory, 2006; Stuss and Alexander, 2000). The question remains, however, whether any connection can be established between such a broad class of cognitive functions and a relatively circumscribed brain location. When it comes to evidence from lesion studies and noninvasive functional neuroimaging, the answer is a qualified "yes."

Functional neuroimaging studies of healthy adults performing executive tasks consistently reported increased activation in prefrontal regions, although involvement of some non-frontal brain areas was also noted. A meta-analysis using activation likelihood estimation (ALE) documented activation of bilateral lateral PFC, anterior cingulate cortex and inferior parietal lobule across studies during WCST (Buchsbaum et al., 2005). Another meta-analysis of fMRI studies revealed reliable activation in the inferior frontal gyrus and anterior cingulate gyrus during Stroop task (Laird et al., 2005). Moreover, a recent meta-analysis of

1653 participants in 189 fMRI samples identified a widespread bilateral fronto-parietal network for working memory (WM) (Rottschy et al., 2012). Specifically, the rostral lateral PFC was related to contrast between WM task and non-WM control, whereas caudal lateral PFC was associated with WM load effects. Thus, the evidence from functional neuroimaging converges on the prefrontal cortex, especially lateral PFC, as a major neural substrate of executive performance.

A recent review and meta-analysis of lesion studies yielded evidence in support of sensitivity if not specificity of executive functions tests to in the frontal lobes (Alvarez and Emory, 2006). In a meta-analysis of 27 lesion studies, an average difference between the patients with frontal lesions and healthy controls on aggregated measures of executive functions was of .83 standard deviation (*sd*), with sample-size-weighted effect of $d = -.78$.

This finding is robust, as the fail-safe analysis indicated that almost three times as many studies with null results would be needed to offset this finding (Alvarez and Emory, 2006). Notably, the difference was smaller in the studies that compared patients with frontal lesions to those who suffered insults in the other regions of the brain ($d = -.57$) but the estimated effect size still indicated certain differential predilection of the executive tests to the frontal lobes. Thus, the evidence that frontal lobe lesions are associated with a large difference in executive functions performance is strong. The greatest impact of the frontal lesions was observed for the Wisconsin Card Sorting Test (WCST, Weigl, 1941) with $d = -.97$ and phonemic word fluency test that yielded an effect of $d = -.80$. Stroop Test of Color-Word Interference (Stroop, 1935) showed only a modest difference of $d = -.30$. Another important moderator examined in that meta-analysis was age, with a curvilinear relationship observed across the examined studies. The scatter plot presented by Alvarez and Emory shows that until the fifth decade, the average age of a sample was unrelated to the effect size, but when the sample consisted of older adults, the difference between lesioned and comparison groups precipitously increased with age. The latter finding is in accord with frequently reported negative associations between age and executive function performance in healthy adults (e.g., Rhodes, 2004).

Although Alvarez and Emory's work documented a significant (albeit not totally unequivocal) connection between the gross integrity of the frontal lobes and executive cognition, their conclusions are limited to lesion studies. Such studies are useful experiments of nature, but in such studies loss of specific functions is inferred from comparison to intact individuals and depends on an assumption of normal performance before the lesion-producing event. Moreover, the effects of lesions, even very circumscribed ones, are not limited to the designated anatomical location due to mass effects and diaschisis, *"...the temporary functional "shock" or deactivation of intact brain regions remote from but connected to the area of primary injury"* (Feeney and Baron, 1986). Thus, reviews of lesion studies and fMRI investigations do not address a question of structure-function relationship in healthy adults, in whom significant individual differences are observed in the cortical size in the absence of frank cortical damage. Although quite a few studies of regional brain volume and thickness are available, it remains unclear whether individual differences in PFC structural characteristics measured in healthy individuals can be tied to individual differences in executive cognition. In other words, in the absence of a clear loss of tissue as

observed in lesion studies, is larger (or thicker) PFC associated with higher scores on measures that purport to assess executive performance or simply put – is bigger better?

A simple log of the extant literature concerned with neuroanatomical correlates of executive cognition in healthy adults reveals a wide range of findings. Some studies show the association between smaller PFC volumes and poor performance on typical executive tasks in normal aging (Gunning-Dixon and Raz, 2003; Raz et al., 1998), others link individual differences in the volume of circumscribed PFC regions to performance on tests of fluid intelligence that are essentially aggregates of executive functions (Colom et al., 2013). In contrast, other studies find no such association (Choi et al., 2004; Morgen et al., 2006; Van Petten et al., 2004), and some even reported negative correlations between PFC volume and executive performance (Duarte et al., 2006; Salat et al., 2002). Thus, the significance of the putative associations between the PFC volume and executive functions can be clarified only by quantitative analyses of the reported effects.

The goal of this meta-analysis was to review the extant studies of the relationships between the measures of prefrontal regional size (volume and cortical thickness) and executive functions in healthy adults and to quantify the reported cognition-structure relationships. We sought to estimate the strength of association between the variation in executive functions and differences in prefrontal size in healthy adults. In addition, we inquired whether the association magnitude was related to sample characteristics (e.g., age), tests applied to evaluation of executive performance, and methods used to measure the target parameters of cortical structure: volume and thickness.

Consistent with the "*bigger is better*" view, we hypothesized that larger PFC volume and greater PFC thickness would be associated with better performance of various tests of executive functions. Particularly, the strongest effect size was expected in lateral PFC rather than orbital and medial PFC, as the functional neuroimaging studies identified lateral PFC to be most relevant to successful executive performance. Because the reviewed studies differed in participants' characteristics and experimental settings, effect sizes were expected to vary across samples. Accordingly, random-effects model was used to compute the summary effects. Though we used random-effects model, the potential heterogeneity across studies was by no means accepted *a priori*. Every estimated effect size was formally tested for heterogeneity. Furthermore, we hypothesized that because of the established relationship between advanced age and deficits in executive functions, a greater mean age and a wider dispersion of age of a sample would be related to larger effect size. Finally, as it was possible that measuring volume in selected regions of interest (ROI) might have missed true effect if the effect was not evenly distributed across the entire ROI, whole-brain voxel-based morphometry (VBM) approach was expected to yield larger effects than ROI approach did. In contrast, we did not expect difference between the effects obtained from cortical thickness studies and those from studies of gray matter volume.

Methods

1. Selection of Studies

A literature search of the computerized PubMed data base was conducted in January, 2014. We used the following search terms: (*frontal* OR *prefrontal*) AND (*volume* OR *volumetric* OR *atrophy* OR *cortical thickness* OR *cortical thinning* OR *morphometry* OR *FreeSurfer*) AND (*Executive* OR *WCST* OR *Wisconsin Card* OR *Stroop* OR *Trail Making Test* OR *TMT* OR *Verbal Fluency* OR *Working Memory*). In addition, relevant references cited in papers found via this search were reviewed.

2. Inclusion Criteria

A search according to the specified terms yielded 965 articles. Titles and abstracts of these articles were examined prior to selection to eliminate studies that clearly did not meet the inclusion criteria. These procedures yielded 231 articles for further examination. The selection of studies was limited to human adults. Case studies were excluded.

Studies were initially selected if they measured prefrontal cortical volume or thickness, included at least one measure of executive function, and contained usable statistics (e.g., correlation) relating prefrontal cortical volume or thickness to executive performance in non-demented human adults. In investigations comparing normal controls with another group of psychiatric or neurological patients, only data from normal control group were considered for inclusion. When two studies emerged from the same lab and there was insufficient information to determine whether the samples were independent, only effect sizes calculated from the study with larger sample size were used. Samples including children that were younger than 18 years were excluded.

Of the 231 articles screened, 52 studies appeared acceptable for inclusion in the meta-analyses. From these 52 studies, two studies (Head et al., 2002; Nestor et al., 2010) were dropped because of significant sample overlap, whereas other two studies with only 26% sample overlap (Gunning-Dixon and Raz, 2003; Raz et al., 1998) were kept. We excluded 17 studies that reported the relationships between executive functions and prefrontal structure size only after controlling some other variables such as age and education, but didn't report zero-order correlations. Therefore, 33 studies were retained in the final selection of meta-analyses (Table 1). One article (Gautam, 2011) examined both cortical thickness and volume. Thus there were 27 studies on prefrontal volume and 7 studies on cortical thickness. Six papers (Burzynska et al., 2012; Gautam et al., 2011; Gianaros et al., 2006; Gur et al., 2000; Kochunov et al., 2009; Takeuchi et al., 2013) examined two separate samples of different age or gender treated as independent sources of the effect size. Final analyses were conducted on 31 independent samples on prefrontal volume, and 10 independent samples on prefrontal gray matter thickness. The total number of participants in the analyzed studies was 3272.

3. Variables Coded

The following variables were gleaned from the selected articles and coded: sample size, age, education, percentage of females in the sample, percentage of left-handers, where the study

was conducted, magnet field strength, software used in imaging processing, cognitive tests used to measure executive function, sub-region of the frontal lobe analyzed in the study, and method of MRI image quantification.

The mean age was treated as a continuous variable, and the standard deviation (*sd*) was selected as an indicator of the age dispersion in a sample. Nakamura and colleagues (Nakamura et al., 2008) reported the average age of 25 healthy subjects, however the relationship between cognitive performance and ROI volume was assessed in 21 subjects. The mean and *sd* of age of 25 subjects were used to estimate the mean and *sd* in 21 subjects, a subsample. Similarly, the standard deviation of age in one study (Salat et al, 2002) was estimated from standard error of age in 31 subjects, though only 30 subjects were included in assessment of the effect size.

The *measurement method* was coded as a binary categorical variable: whole-brain evaluation vs. ROI-based morphometry. We did not distinguish among various type and brands of software employed in these analyses.

Cognitive test type was coded into several categories: Wisconsin Card Sorting Test (WCST), Trail Making Test (TMT), fluency, working memory (WM), interference and composite score. If the results of two or more tests of executive functions were reported for a sample, then a combined effect size was calculated by averaging the Fisher's *z*-scores of these tests, and was used as the overall effect size of this sample.

The PFC *region* examined in each study was coded as whole prefrontal cortex (PFC), lateral prefrontal (LPFC), orbito-frontal (OFC) and medial prefrontal (MPFC) cortices. The overall effect size for each sample was assessed from the whole prefrontal area, if it is available. Otherwise, if the associations between executive functions and prefrontal structure sizes were reported in two or more sub-regions of prefrontal structure, then the effects were averaged into one overall effect size.

The other coded variables, including education, percentage of females, percentage of left-handers, place where the study was conducted, magnetic strength and software used in imaging processing, were not included in the moderator analyses due to insufficient reports or minimal variability.

For the majority of cognitive measures, higher scores were indicative of better performance. However, for some measures such as reaction time and number of errors, a higher score reflected poorer performance. Therefore, these measures were reversely coded so that all outcome measures were in the same direction, with higher score indicating better performance. Accordingly, positive effect size implied that larger prefrontal cortex was associated with better executive functions, and negative effect size indicated that larger prefrontal cortex was associated with poorer executive performance.

4. Computation of Effect Sizes and Outlier Analysis

All study results were converted to correlations (*r*) and Fisher transformed *z*-scores (Z_r). Z_r was used as effect size. The computations were based on correlations (*r*), *p*-values, and

bivariate linear regression statistics (R^2). Only raw correlations (bivariate correlations) were used to compute effect sizes. The majority of studies reported Pearson correlations r, whereas three studies (Baare et al., 1999; Morgen et al., 2006; Zuffante et al., 2001) reported Spearman rank-order correlations r. These Spearman rank-order correlations were treated the same as Pearson correlations. For studies that reported a "significant" relationship, but did not report an actual statistic, the effect sizes were conservatively estimated from $p = 0.049$ if alpha level of 0.05 was used, or $p = 0.009$ if alpha level of 0.01 was used in the study. For studies that reported a "non-significant" relationship, but did not report an actual statistic, the effect sizes were conservatively estimated from $r = 0$. One article (Tullberg et al., 2004) reported bivariate linear regression statistics (R^2) for the association between executive functions and volume of frontal gray matter. The correlation was estimated from square root of R^2.

For studies that used a whole-brain approach to assess association between executive functions and structure size, the average effect in a cluster was used if it is available. Otherwise, effect size was conservatively estimated from p value. The peak values were not used.

The indices of brain-cognition association in two hemispheres were combined. When results were reported separately for right and left hemispheres, the arithmetic averages of effects were computed. If an article reported significance in one hemisphere, but did not mention the other hemisphere, then it was presumed that the other hemisphere was not reported because of non-significance, and thus effect size was estimated as $r = 0$.

If both the volumes of frontal lobe (gray matter + white matter) and prefrontal gray matter were reported in a paper, the correlation between executive functions and gray matter volume was selected. Analyses were conducted both with and without the studies reporting only the correlation between executive functions and volume of gray matter plus white matter.

For each executive task, if a study used more than one measure, e.g. perseverative errors and categories achieved in WCST, these measures were averaged to generate an overall effect size for the test. If more than one test of executive functions or more than one region was reported in a study, then effect sizes were collapsed so that there was one single effect size for each sample. This single effect size was calculated by averaging the Fisher's z-scores across the executive tests and averaging the Fisher's z-scores across the sub-regions. All standardized effect sizes (Z_r) were assessed for outliers.

5. Statistical methods

The *metafor* package in R was used to conduct the meta-analyses with random-effects model estimation. The meta-analyses followed the DerSimonian-Laird method (DerSimonian and Laird, 1986), and the underlying formulas are listed in Appendix A. The summary effect sizes of relationship between executive functions and prefrontal cortex size were separately analyzed for cortical thickness and volume. Random-effects model was used because there might be different effect sizes underlying different studies included. The studies differed in participants and in methods, and as a result, there was no reason to assume that the true

effect size was exactly the same in all the studies. Thus to capture the variability in true effect size, we chose random-effects model rather than fixed-effect model.

Moderating effects of *sample mean age, sample age dispersion (standard deviation, sd) and measurement method* were estimated separately. *Cognitive test type* and *sub-region* were not used as moderators because a study could employ more than one cognitive test and might assess the cognition-structure association in more than one sub-region. Instead, in subsidiary analysis exploring the potential effect of *region*, the summary effect sizes were computed individually for the volumes of LPFC, OFC and MPFC, without collapsing effect sizes across subregions. Similarly, in another subsidiary analysis exploring the potential effect of *cognitive test type*, the summary effect sizes were calculated individually for WCST, TMT, WM, fluency test and interference test.

Results

1. Strength of association between executive functions and PFC volume

The mean of Fisher's z-transformed correlation between assorted measures of executive functions and prefrontal volume across 31 samples was 0.153 (*se* = 0.039, Z_{test} = 3.927, *p* <. 001, 95% CI: 0.077/0.230; Figure 1), indicating a moderate positive association between executive performance and the volume of PFC. The effect size corresponds to Cohen's *d* = . 31.

The fail-safe analysis (Orwin, 1983) revealed that 17 additional studies with null findings would be needed to offset the observed effect size. The estimate did not change significantly after excluding six samples that reported correlations between executive functions and combined volume of frontal gray matter and white matter. In the remaining 25 samples that measured prefrontal gray matter volume, the estimated mean effect size was 0.167 (*se* = 0.045, Z_{test} = 3.702, *p* <.001, 95% CI: 0.078/0.255), corresponding to Cohen's *d* = .34. The heterogeneity of the summary effect was significant, before ($Q \chi^2$ (30) = 118.03, *p* < .001) and after excluding the samples of combined frontal gray and white volume ($Q \chi^2$ (24) = 114.46, *p* < .001).

2. Summary effect for executive functions-PFC thickness studies

The weighted mean effect size, correlation between executive functions and prefrontal cortical thickness across ten samples was 0.096 (*se* = 0.042, Z_{test} = 2.254, *p* = 0.024, 95% CI: 0.013/0.179; Figure 2), corresponding to Cohen's *d* = .19 and indicating a weak but significant association between cognitive and neuroanatomical indices. The heterogeneity among these 10 samples showed a trend towards significance ($Q \chi^2$ (9) = 14.98, *p* = 0.091).

Better performance on the tests of executive functions was associated with larger prefrontal cortical volume and cortical thickness, as indicated by the significantly positive mean effect sizes. Although the mean effect size for volume was greater than that for cortical thickness, the 95% confidence intervals of effect sizes overlapped. Thus, the two indices of prefrontal cortical integrity exhibited equivalent modest association with performance on tests of executive functions.

3. Heterogeneity and moderator tests

Strength of association between executive functions and measures of structural integrity of the PFC, varied significantly across samples. Fisher's *z* transformed correlation coefficient ranged from −0.25 to 0.56 (Figure 1 and Figure 2).

The random-effects model used in the current meta-analysis is predicated on the idea that the effect sizes are sampled from a distribution, and that the true effects do not have to be equal across studies. Thus, under the random-effects model, a part of the dispersion in observed effects stems from a sampling error, whereas the other part may reflect real differences in effect size across studies. In a sample of 31 studies of PFC volume and executive functions, 75% of the total variability in the effect size was attributable to heterogeneity, and the estimated standard deviation of the distribution of the true effects (τ) was 0.173 (Figure 3). For 10 samples that had cortical thickness measures, 40% of the total variability was due to heterogeneity, and the estimated τ was 0.077 (Figure 3).

The observed heterogeneity suggested that some factors might affect the relationship between executive performance and prefrontal volume or cortical thickness. However, because the number of cortical thickness studies (10) was too small for moderator analyses, we evaluated moderator effects only for 31 samples of prefrontal volume. In that sample of studies, we examined three moderators: the sample *mean age*, sample age *sd* and the *measurement method* used for assessment of PFC volume.

3.1. Parameters of age distribution as moderators of effect size—The *mean age* of a sample did not moderate the effect size: $Q_M \chi^2 (1) = 0.029$, $p = 0.866$, figure 4. Differences in the *mean age* among samples did not significantly account for the heterogeneity of correlation between prefrontal volume and executive functions. As presented in Table 2, the 95% confidence intervals of estimated mean effects overlapped across age groups.

On the other hand, the dispersion of age in the sample, *age sd*, emerged as a significant moderator of the effect size: $Q_M \chi^2 (1) = 17.097$, $p < 0.001$. As illustrated in Figure 5, larger *age sd* was associated with stronger correlation between PFC volume and executive functioning. However, after accounting for the contribution of sample age variance to the estimated effect size, the residual heterogeneity remained significant: $\chi^2 (27) = 62.793$, $p < 0.001$. Thus, additional factors influenced the relationship between executive functions and PFC volume.

3.2. Volume measurement method as a moderator—The *measurement method* significantly moderated correlation effects ($Q_M \chi^2 (1) = 4.076$, $p = 0.044$), although the 95% confidence intervals of estimated mean effects in two subgroups of studies overlapped: 0.116/0.261 across 26 samples using ROI approach and −0.225/0.227 across 5 samples using whole-brain approach. Thus, in contrast to studies using whole-brain method of local volume estimation, the studies employing ROI measures reported stronger association between prefrontal cortex volume and executive functions.

4. Effect of test type

In addition to *mean age, variance of age* and *measurement method*, two factors were evaluated as potential variables influencing effect sizes: the type of cognitive tests to assess executive functions, and sub-region of prefrontal area correlated with test performance. They were not examined as moderators, because a number of studies used more than one type of cognitive test to evaluate the correlation between prefrontal volume and executive performance, and several studies reported effects in multiple sub-regions of prefrontal cortex. Instead, the summary effects were individually estimated at each level of these two variables.

The mean effect sizes were calculated for each of the tests in the study, including Wisconsin Card Sorting Test (WCST), Trail Making Test (TMT), working memory (WM), fluency test and interference test, and are summarized in Table 3. As presented in Table 3, heterogeneity was significant for TMT and WM but not for WCST, fluency and interference tests. Further examination revealed, however, that the heterogeneity in TMT could be explained by an outlier (Paul et al., 2009, effect size = −0.321, Z-score = −2.109). After removal of this outlier, the heterogeneity in TMT was no longer significant ($p = 0.458$). On the other hand, no outliers were detected in the distribution of WM effects. Moreover, moderator analysis indicated that *mean age* and *measurement method* did not moderate the heterogeneity of effects in WM, whereas the age dispersion moderated the correlation between working memory and PFC volume ($Q_M \chi^2 (1) = 8.102$, $p = 0.004$). Among the 16 samples reporting effect sizes of WM, 6 used digit backwards span (DBS) to measure WM; another 5 reported the correlation between PFC volume and composite WM scores; 2 samples employed a four word short-term memory test; 2 studies used CANTAB and the other one used spatial delayed response task and self-ordered pointing test to assess WM ability. The heterogeneity test was not significant in the samples of digit backwards span, suggesting that the observed heterogeneity in WM effects might result from the vast diversity of WM tests. The estimated τs were 0.091, 0, 0.049, 0, 0.147, 0, respectively, for WCST, fluency test, interference test, TMT without outlier, WM, and DBS. As the estimated τs equaled zero for fluency test, TMT and DBS, the estimations of these three tests collapsed into fixed-effect model. The estimated distributions of true effect sizes of each cognitive test are presented in Figure 3.

The estimated mean effects for WCST (Figure 6), WM (Figure 7), DBS and interference test (Figure 8) differed significantly from zero (all $p < 0.001$; Table 3), whereas the estimated mean effect for fluency test (Figure 9) did not ($p = 0.207$). Without the identified outlier, the estimated mean effect in TMT was significant and positive ($p < 0.001$), but was not significantly different from zero if the outlier was kept (Figure 10).

The positive correlation of prefrontal volume with WCST score was significantly larger than its correlations with TMT, DBS and fluency scores, as indicated by the non-overlapping 95% confidence intervals (Table 3) and significant differences in estimated mean effects (WCST-TMT: $Z^*_{diff} = 3.598$; WCST-Fluency: $Z^*_{diff} = 3.644$; WCST-DBS: $Z^*_{diff} = 2.630$; all $p < 0.01$). Although the mean effect size for interference tests was larger than for TMT ($Z^*_{diff} = 3.609$, $p < 0.001$), fluency tests ($Z^*_{diff} = 3.630$, $p < 0.001$) and DBS ($Z^*_{diff} = 2.596$, $p < 0.01$), the small number of samples that used interference tests (5) casts some doubt on

this finding. No significant difference was found among the mean effect sizes of DBS, TMT, and fluency tests.

5. Effect size in PFC sub-regions

The effect sizes were examined separately in the three sub-regions of prefrontal cortex: LPFC (Figure 11), OFC (Figure 12) and MPFC (Figure 13). As summarized in Table 4, the heterogeneity tests were significant in both LPFC and OFC ($p < 0.001$), and showed a trend towards significance in MPFC ($p = 0.089$). The estimated τs were 0.171, 0.246 and 0.141 for LPFC, OFC and MPFC, respectively. The effect sizes in LPFC and OFC were significantly moderated by dispersion of sample age ($Q_M \chi^2 (1) = 9.087$ for LPFC and $Q_M \chi^2 (1) = 10.452$ for OFC, both $p < 0.01$). After controlling for sample age *sd*, the residual heterogeneity was no longer significant for OFC ($p = 0.14$) but still significant for LPFC ($Q_E \chi^2 (16) = 34.532$, $p < 0.01$). There was a significantly positive correlation between executive functions and LPFC, MPFC volume (both $p < 0.05$), but not in OFC. The estimated mean effect in OFC was smaller than in LPFC ($Z^*_{diff} = 2.391$, $p = 0.017$), and with a trend to be smaller than in MPFC ($Z^*_{diff} = 1.893$, $p = 0.058$). The fail-safe analysis revealed that 23 additional studies with null findings (i.e., more than doubling of the sample size) would be needed to offset the observed effect size in LPFC.

Discussion

The main finding of this meta-analysis is that in healthy adults, larger prefrontal cortex, especially its lateral part, is associated with better performance on tests of executive functions. Although the association between the analyzed measures of brain and cognition is modest and the range of reported effects is very broad, the link is quite robust in lateral PFC area. This finding is consistent with the reported associations between PFC lesions and performance on executive tasks (Alvarez and Emory, 2006), although the magnitude of the association is not as strong.

Significant heterogeneity of effect sizes indicates that multiple factors may account for the reported differences in findings, but in this analysis we were able to examine only a few of potential moderators. A notable finding here is that the more age-heterogeneous the study sample, the more likely is that study to report a significant association between PFC volume and executive performance. Thus, within-sample age variability rather than differences in the mean age of the sample contribute to the effect size. The importance of within-sample age variability for the link between PFC volume and executive performance may reflect a fact that both PFC (Raz and Kennedy, 2009) and executive functions (Rhodes, 2004; Rodriguez-Aranda and Martinussen, 2006; Salthouse, 1991; Tombaugh, 2004) evidence significant age differences. Thus, differences in age-related variability may be a reason for discrepancy in the size and direction of effects reported in the studies. The variance of executive functioning and PFC volume could be limited if the age range was compressed, and the association between cognitive and brain variable could be reduced even further if both distributions were truncated. It appears, therefore, that the association of executive test scores with the PFC volume may reflect significant commonality of the variance of these variables with age. This problem has been pointed out in several analyses of mediational

models in aging and development (Hofer and Sliwinski, 2001; Lindenberger and Potter, 1998) and needs to be addressed within the framework of longitudinal design (Raz and Lindenberger, 2011).

Additional factors, some of them age-related may moderate the relationship between PFC size and executive cognition. Unfortunately, studies vary in the level of detail in their sample description. Although all report age range and variance, information regarding other factors that may affect both brain structure and cognitive performance is frequently left out. Several age-related health factors, especially vascular risk indicators may exert negative influence on both PFC (Jennings et al., 2012; Raz et al., 2003) and executive test scores (Raz et al., 2003; Saxby et al., 2003; Vicario et al., 2005). As noted more than six decades ago with reference to such effects, "*impairment of cerebral functions equivalent to that seen in patients with surgical removal of both frontal lobes may occur early in the course of essential hypertension*" (Apter et al., 1951). Failure to include measures of vascular health in the models assessing the associations between structural characteristics of the PFC and executive performance needs to be addressed before we are able to fully describe the nature of the observed structure-function associations.

The observed relationship between the PFC volume and executive performance was stronger in the studies that used ROI-based methods than in those that relied on whole-brain computational methods. That difference may reflect methodological peculiarities of whole-brain measurement, in which voxels within anatomical structures rather than valid anatomical structures such as gyri serve as units of analysis. With regards to age-related differences, correspondence between voxel-oriented (peak of significance) regional volumes and ROI-based measures is weak (Kennedy et al., 2009). Multiple comparisons correction across a large number of voxels may be too conservative. As a result, the overall effect generated by whole-brain analysis was smaller than the overall effect from ROI measurement.

Not all tests of executive functions reveal similar associations with PFC volume. Even within the limitations of the small sample of studies, the WCST yields significantly stronger associations than other measures do, and some – verbal fluency – shows no link to PFC volume at all. The literature on WCST, a test that was introduced more than 70 years ago (Weigl, 1941) is rich, and many aspects of this complex test have been examined in clinical and experimental studies. Its sensitivity (though not specificity) to frontal lesions has been confirmed (Alvarez and Emory, 2006; Demakis, 2003) and with the addition of this finding, it seems that investigations of the neural foundations of executive functions may benefit from deconstruction of the WCST and identifying the critical ingredients that are responsible for its sensitivity to normal variations of the PFC volume. Some studies suggest working memory as the key to age differences in WCST (Fristoe et al., 1997; Hartman et al., 2001), but as this meta-analysis shows, use of WM tests yields smaller effect than the use of WCST. The latter finding may, however, reflect the heterogeneity of WM tests employed by various studies.

Unlike other executive tests in this sample of studies, the WM measures exhibited significant within-group heterogeneity of effect sizes, i.e. it mattered what type of WM test

was used to asses executive functions. It is possible that the construct of "working memory" may be not sufficiently defined for the purpose of examining associations with regional brain volumes. Moreover, functional neuroimaging literature shows that WM is supported by a wide variety of cross-cortical networks that include PFC but are not limited to it (Cabeza and Nyberg, 2000; Owen et al., 2005; Yendiki et al., 2010). Comparison of effect sizes in the regions that participate in such networks as well as in control regions that are not expected to influence executive processing would be of great interest. Dissociations among these regions would bolster the claim of specificity of PFC role in executive performance; whereas finding of similar effect sizes would negate such a proposition. Unfortunately, only a handful of the reviewed studies have attempted such dissociations, and there is little consistency among them in selection of control regions.

Primary visual cortex (VC) would be a good choice to play a role of a control area, because according to the literature, it is supposed to have little involvement in the executive tasks (Laird et al., 2005; Rottschy et al., 2012). Regrettably, only two studies compared the association between executive function and regional brain volumes of PFC and primary VC. In both, better working memory was associated with larger dorsolateral PFC volume, whereas the correlation with volume of occipital cortex is not significant (Raz et al., 1999; unpublished data from Head et al., 2009). In other studies, secondary visual cortex (fusiform gyrus, FG) alone or in combination with VC exhibited correlations with executive functions that were smaller (though not significantly) than those observed for the LPFC (Raz et al., 1998; Gunning-Dixon & Raz, 2003). Another study using occipital gray matter volume as control structure found significant correlation with executive functions neither for occipital cortex, nor for the PFC volume (Van Petten et al., 2004). Although the differences between correlations did not reached significance in some of these studies, in almost all of them the associations between executive tests and LPFC were stronger than between the same tests and volumes of primary and secondary visual cortices. Thus, we surmise that more numerous comparison studies could have made a stronger case for specificity of the LPFC-executive functions relationship. As summarized in Table 5, positive and null effects were reported in a wide range of non-PFC regions, though no negative relationships ("smaller – better") were reported in the reviewed ROI-based studies.

A limited remedy to the lack of ROI comparison studies may be found in comparing the regions within the PFC. As the comparison among PFC sub-regions showed, performance on executive functions tests was significantly related to the volume of lateral and medial PFC but not of the OFC. The null effect in orbital PFC suggests that the positive effect observed in PFC, especially lateral PFC, cannot reflect differences across the brain. This regional comparison supports the specific role of lateral PFC in executive functioning.

The problem of establishing specificity of a region for support of a cognitive operation is seemingly easier to find in VBM studies that examine the associations of executive performance with local brain density and volume that are hypotheses-free and relay solely on statistical decision criteria. Such studies are expected to reveal a larger set of structure-function relationships than hypotheses-driven ROI studies can. Indeed, voxel-based studies revealed associations between executive functions and local cortical volume or thickness in PFC as well as non-PFC regions. The latter included paracentral lobule (Burzynska et al.,

2012; Haldane et al., 2008), parietal cortex (Burzynska et al., 2012; Dickerson et al., 2008; Haldane et al., 2008; Tu et al., 2012), cingulate gyrus (Haldane et al., 2008) and cerebellum (Haldane et al., 2008), which all evidenced positive correlations between local volume or thickness and executive test scores. In addition, however, VBM studies yielded negative associations between executive cognition and the size of temporal (Burzynska et al., 2012) and insular cortices (Koutsouleris et al., 2010), i.e. *greater* local volume or thickness was associated with *poorer* performance. Because there was little consistency among VBM studies in reporting the associations between executive tests and non-PFC regions we could not subject these findings to meta-analysis. Notably, none of the VBM studies reported significant correlations between executive functions and occipital cortex volume or density thus presenting a good reason to view that part of the brain as a suitable control region.

The results of this meta-analysis do not yield themselves to a simple interpretation, and it is unclear why bigger brain regions are associated with better cognitive performance. The neurobiological reality behind variability in volumes and cortical thickness observed on MRI is complicated and poorly understood. Although, recent work in rodents points to a strong link between changes in local volume and density of neuropil (Qiu et al., 2013), other studies reveal association between hippocampal volume and attrition of neurons (Bobinski et al, 2000; Lee et al., 1995). Other possibilities, such as reduction in neuronal size, decline in glial population, and reduction in capillary density remain unexplored. Because cognitive training may affect local brain structure (e.g., Engvig et al., 2010), it is possible that differences in any of the brain components listed above could be reciprocally related to cognitive performance, either by providing better infrastructure for neurocomputation or by representing a response to increased cognitive activity. Because the cause and effect relationship between local cortical volumes and executive cognitions cannot be established in cross-sectional studies, in this meta-analysis, we take a correlational view and evaluate the strength of associations between regional structural measures and executive functions.

Limitations of the current meta-analyses

The conclusions of this meta-analysis are constrained by several limitations. First, we aggregated the findings across variable definitions of the PFC regions across the studies. Some investigators measured relatively large PFC regions incorporating many Brodmann areas (8, 9, 10, 45, 46) to ensure high reliability (Gautam, 2011; Hanninen et al., 1997; Raz et al., 1998). In other studies, however, the volume of much smaller regions of the PFC was used in investigation of its association with executive functions. For example, Zuffante et al. manually traced Brodmann's area 46 (Zuffante et al., 2001); whereas in another study, the PFC ROI comprised areas 9 and 46 (Paul et al., 2009). Although there is a basic agreement on what part of the frontal lobe constitutes PFC (Fuster, 2008), each laboratory uses locally developed and validated rules to define PFC ROIs. Not all studies report the reliability of the measures and those that do, use various indices of reliability. Thus, discrepancy in reliability and validity of PFC measures may account for a significant part of inter-study variance in effect size.

The problem of heterogeneity of ROI definitions is not restricted to manual morphometry studies. The regions demarcated by whole-brain semi-automatic methods, also differ among

studies. The prefrontal cortex is a cytoarchitectonically and functionally heterogeneous region (Petrides and Pandya, 2002; Stuss et al., 2002). Anatomically, distinct regions in PFC project to multiple cortical regions via white matter pathways that may be differentially related to cognitive performance (Madden et al., 2009 for review). Moreover, PFC has dense connections to the subcortical nuclei and the cerebellum (Strick et al., 2009). Various executive functions may depend on different regions of prefrontal cortex (Stuss et al., 2002; Stuss et al., 1995). However, variability on ROI definitions and cortical parcellation among the studies may not be greater than variability in the size and location of the damage in lesion studies (Alvarez and Emory, 2006). Thus, although heterogeneity of anatomical measures limits the validity of this meta-analysis, it should not preclude integration across studies. If anything, refinement of the measures may reduce error variance and increase the observed effect size in the future studies.

A common limitation of many meta-analyses is publication bias or a file-drawer effect. Indeed, the effect sizes in published studies are generally larger than the effect sizes in unpublished studies (Lipsey and Wilson, 1993), and the overall effect size may be overestimated. In this meta-analysis, we have shown that at least for the association between executive performance and the lateral PFC volume, the effect is robust, even with potential overestimation taken into account. To nullify the observed overall effect in lateral PFC, the proverbial file drawer will have to contain at least 23 studies with null finding, more than 120% of valid samples used in the current meta-analysis. This minimum number will even increase, if studies with positive effect size are also included in the file drawer.

In conclusion, a quantitative analysis of the extant studies revealed a modest but robust positive association between performance on tests of executive functions and prefrontal cortical size in healthy adults. Larger PFC, especially lateral and medial PFC volume and larger prefrontal cortical thickness were associated with better executive performance, although such relationship is stronger for some tests and in samples with greater variation of age. In short, the results support at least a limited version of the "*bigger is better*" hypothesis of brain-behavior relationships. Improvement in specificity of executive tests and refinement of neuroanatomical measurements as well as rigorous description of the sample health characteristics should help to improve the quantification of structure-function relationship in the future studies.

Acknowledgments

Appendix A. Underlying formulas of the current meta-analyses

1. Converting correlation coefficient (*r*) to Fisher transformed *z*-score (Z_r):

$$Z_r = 0.5 \ln \frac{1+r}{1-r}.$$

2. Generic summary effect $M = \dfrac{\sum Z_r (n-3)}{\sum (n-3)}$, where *n* is the sample size of each sample.

3.
   Weighted sum of squares: $Q = \sum\limits_{i=1}^{k} W_i(Y_i - M)^2$, where $W_i = n-3$, Y_i is the effect size (Fisher's z) of each sample and k is the number of samples.

4.
   $$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

5.
   Variance of the true effects: $T^2 = \dfrac{Q - df}{C}$, where $df = k - 1$.

6.
   Within-study variance: $V_{Y_i} = \dfrac{1}{n-3}$.

7.
   Weight of each study under random-effects model: $W_i^* = \dfrac{1}{V_{Y_i} + T^2}$.

8.
   Weighted mean under random-effects model: $M^* = \dfrac{\sum W_i^* Y_i}{\sum W_i^*}$.

9.
   Variance of the weighted mean: $V_{M^*} = \dfrac{1}{\sum W_i^*}$.

10.
   Standard error of the weighted mean: $SE_{M^*} = \sqrt{V_{M^*}}$.

11. The 95% confidence upper and lower limits of weighted mean:

   a. $UL_M{}^* = M^* + 1.96 \times SE_M{}^*$,

   b. $LL_M{}^* = M^* - 1.96 \times SE_M{}^*$.

12.
   Converting Fisher's z to correlation coefficient: $r = \dfrac{e^{2z} - 1}{e^{2z} + 1}$.

13.
   Evaluating the significance of r by: $Z^* = \dfrac{M^*}{SE_{M^*}}$.

14. The proportion of real variance of effect size in the total observed dispersion:

   $I^2 = \dfrac{Q^2 - df}{Q^2}$.

15.
   Comparing the mean effect for subgroups: $Z^*{}_{diff} = \dfrac{M_B^* - M_A^*}{\sqrt{V_{M_A^*} + V_{M_B^*}}}$.

References

Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. Neuropsychol Rev. 2006; 16:17–42. [PubMed: 16794878]

Apter NS, Halstead WC, Heimburger RF. Impaired cerebral functions in essential hypertension. Am J Psychiatry. 1951; 107:808–813. [PubMed: 14829603]

Baare WF, Hulshoff Pol HE, Hijman R, Mali WP, Viergever MA, Kahn RS. Volumetric analysis of frontal lobe regions in schizophrenia: relation to cognitive function and symptomatology. Biol Psychiatry. 1999; 45:1597–1605. [PubMed: 10376121]

Baddeley A. Exploring the Central Executive. The Quarterly journal of experimental psychology. A, Human experimental psychology. 1996; 49:5–28.

Bender AR, Raz N. Age-related differences in memory and executive functions in healthy APOE varepsilon4 carriers: the contribution of individual differences in prefrontal volumes and systolic blood pressure. Neuropsychologia. 2012; 50:704–714. [PubMed: 22245009]

Bobinski M, de Leon MJ, Wegiel J, Desanti S, Convit A, Saint Louis LA, Rusinek H, Wisniewski HM. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. Neuroscience. 2000; 95:721–725. [PubMed: 10670438]

Buchsbaum BR, Greer S, Chang WL, Berman KF. Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. Hum Brain Mapp. 2005; 25:35–45. [PubMed: 15846821]

Burzynska AZ, Nagel IE, Preuschhof C, Gluth S, Backman L, Li SC, Lindenberger U, Heekeren HR. Cortical thickness is linked to executive functioning in adulthood and aging. Hum Brain Mapp. 2012; 33:1607–1620. [PubMed: 21739526]

Cabeza R, Nyberg L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. J Cogn Neurosci. 2000; 12:1–47. [PubMed: 10769304]

Chan RC, Shum D, Toulopoulou T, Chen EY. Assessment of executive functions: review of instruments and identification of critical issues. Arch Clin Neuropsychol. 2008; 23:201–216. [PubMed: 18096360]

Cheung AM, Mitsis EM, Halperin JM. The relationship of behavioral inhibition to executive functions in young adults. J Clin Exp Neuropsychol. 2004; 26:393–404. [PubMed: 15512928]

Choi JS, Kang DH, Kim JJ, Ha TH, Lee JM, Youn T, Kim IY, Kim SI, Kwon JS. Left anterior subregion of orbitofrontal cortex volume reduction and impaired organizational strategies in obsessive-compulsive disorder. J Psychiatr Res. 2004; 38:193–199. [PubMed: 14757334]

Colom R, Burgaleta M, Roman FJ, Karama S, Alvarez-Linera J, Abad FJ, Martinez K, Quiroga MA, Haier RJ. Neuroanatomic overlap between intelligence and cognitive factors: Morphometry methods provide support for the key role of the frontal lobes. Neuroimage. 2013; 72:143–152. [PubMed: 23357078]

Demakis GJ. A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. Neuropsychology. 2003; 17:255–264. [PubMed: 12803431]

DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–188. [PubMed: 3802833]

Dickerson BC, Fenstermacher E, Salat DH, Wolk DA, Maguire RP, Desikan R, Pacheco J, Quinn BT, Van der Kouwe A, Greve DN, Blacker D, Albert MS, Killiany RJ, Fischl B. Detection of cortical thickness correlates of cognitive performance: Reliability across MRI scan sessions, scanners, and field strengths. Neuroimage. 2008; 39:10–18. [PubMed: 17942325]

Duarte A, Hayasaka S, Du A, Schuff N, Jahng GH, Kramer J, Miller B, Weiner M. Volumetric correlates of memory and executive function in normal elderly, mild cognitive impairment and Alzheimer's disease. Neurosci Lett. 2006; 406:60–65. [PubMed: 16904823]

Engvig A, Fjell AM, Westlye LT, Moberget T, Sundseth O, Larsen VA, Walhovd KB. Effects of memory training on cortical thickness in the elderly. Neuroimage. 2010; 52:1667–1676. [PubMed: 20580844]

Feeney DM, Baron JC. Diaschisis. Stroke. 1986; 17:817–830. [PubMed: 3532434]

Fristoe NM, Salthouse TA, Woodard JL. Examination of age-related deficits on the Wisconsin Card Sorting Test. Neuropsychology. 1997; 11:428–436. [PubMed: 9223147]

Fuster, JM. The Prefrontal Cortex. 4th ed.. Boston: Academic Press; 2008.

Garlinghouse MA, Roth RM, Isquith PK, Flashman LA, Saykin AJ. Subjective rating of working memory is associated with frontal lobe volume in schizophrenia. Schizophr Res. 2010; 120:71–75. [PubMed: 20303715]

Gautam P. Relationships between cognitive function and frontal grey matter volumes and thickness in middle aged and early old-aged adults: The PATH Through Life Study. NeuroImage (Orlando, Fla.). 2011; 55:845–855.

Gianaros PJ, Greer PJ, Ryan CM, Jennings JR. Higher blood pressure predicts lower regional grey matter volume: Consequences on short-term information processing. Neuroimage. 2006; 31:754–765. [PubMed: 16488626]

Goldstein KE, Hazlett EA, Savage KR, Berlin HA, Hamilton HK, Zelmanova Y, Look AE, Koenigsberg HW, Mitsis EM, Tang CY, McNamara M, Siever LJ, Cohen BH, New AS. Dorso- and ventro-lateral prefrontal volume and spatial working memory in schizotypal personality disorder. Behav Brain Res. 2011; 218:335–340. [PubMed: 21115066]

Grambaite R, Selnes P, Reinvang I, Aarsland D, Hessen E, Gjerstad L, Fladby T. Executive dysfunction in mild cognitive impairment is associated with changes in frontal and cingulate white matter tracts. J Alzheimers Dis. 2011; 27:453–462. [PubMed: 21841261]

Gunning-Dixon FM, Raz N. Neuroanatomical correlates of selected executive functions in middle-aged and older adults: a prospective MRI study. Neuropsychologia. 2003; 41:1929–1941. [PubMed: 14572526]

Gur RE, Cowell PE, Latshaw A, Turetsky BI, Grossman RI, Arnold SE, Bilker WB, Gur RC. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. Arch Gen Psychiatry. 2000; 57:761–768. [PubMed: 10920464]

Haldane M, Cunningham G, Androutsos C, Frangou S. Structural brain correlates of response inhibition in Bipolar Disorder I. J Psychopharmacol. 2008; 22:138–143. [PubMed: 18308812]

Hanninen T, Hallikainen M, Koivisto K, Partanen K, Laakso MP, Riekkinen PJ Sr, Soininen H. Decline of frontal lobe functions in subjects with age-associated memory impairment. Neurology. 1997; 48:148–153. [PubMed: 9008510]

Hartberg CB, Sundet K, Rimol LM, Haukvik UK, Lange EH, Nesvag R, Dale AM, Melle I, Andreassen OA, Agartz I. Brain cortical thickness and surface area correlates of neurocognitive performance in patients with schizophrenia, bipolar disorder, and healthy adults. J Int Neuropsychol Soc. 2011; 17:1080–1093. [PubMed: 22013998]

Hartman M, Bolton E, Fehnel SE. Accounting for age differences on the Wisconsin Card Sorting Test: decreased working memory, not inflexibility. Psychol Aging. 2001; 16:385–399. [PubMed: 11554518]

Head D, Kennedy KM, Rodrigue KM, Raz N. Age differences in perseveration: cognitive and neuroanatomical mediators of performance on the Wisconsin Card Sorting Test. Neuropsychologia. 2009; 47:1200–1203. [PubMed: 19166863]

Head D, Raz N, Gunning-Dixon F, Williamson A, Acker JD. Age-related differences in the course of cognitive skill acquisition: the role of regional cortical shrinkage and cognitive resources. Psychol Aging. 2002; 17:72–84. [PubMed: 11931289]

Hofer SM, Sliwinski MJ. Understanding ageing - An evaluation of research designs for assessing the interdependence of ageing-related changes. Gerontology. 2001; 47:341–352. [PubMed: 11721149]

Jennings JR, Mendelson DN, Muldoon MF, Ryan CM, Gianaros PJ, Raz N, Aizenstein H. Regional grey matter shrinks in hypertensive individuals despite successful lowering of blood pressure. J Hum Hypertens. 2012; 26:295–305. [PubMed: 21490622]

Kennedy KM, Erickson KI, Rodrigue KM, Voss MW, Colcombe SJ, Kramer AF, Acker JD, Raz N. Age-related differences in regional brain volumes: A comparison of optimized voxel-based morphometry to manual volumetry. Neurobiology of Aging. 2009; 30:1657–1676. [PubMed: 18276037]

Kochunov P, Robin DA, Royall DR, Coyle T, Lancaster J, Kochunov V, Schlosser AE, Fox PT. Can structural MRI indices of cerebral integrity track cognitive trends in executive control function during normal maturation and adulthood? Hum Brain Mapp. 2009; 30:2581–2594. [PubMed: 19067326]

Koutsouleris N, Patschurek-Kliche K, Scheuerecker J, Decker P, Bottlender R, Schmitt G, Rujescu D, Giegling I, Gaser C, Reiser M, Moller HJ, Meisenzahl EM. Neuroanatomical correlates of executive dysfunction in the at-risk mental state for psychosis. Schizophr Res. 2010; 123:160–174. [PubMed: 20826078]

Kozicky JM, Ha TH, Torres IJ, Bond DJ, Honer WG, Lam RW, Yatham LN. Relationship between frontostriatal morphology and executive function deficits in bipolar I disorder following a first manic episode: data from the Systematic Treatment Optimization Program for Early Mania (STOP-EM). Bipolar Disord. 2013; 15:657–668. [PubMed: 23919287]

Laird AR, McMillan KM, Lancaster JL, Kochunov P, Turkeltaub PE, Pardo JV, Fox PT. A comparison of label-based review and ALE meta-analysis in the Stroop task. Hum Brain Mapp. 2005; 25:6–21. [PubMed: 15846823]

Lee N, Tien RD, Lewis DV, Friedman AH, Felsberg GJ, Crain B, Hulette C, Osumi AK, Smith JS, VanLandingham KE, et al. Fast spin-echo, magnetic resonance imaging-measured hippocampal volume: correlation with neuronal density in anterior temporal lobectomy patients. Epilepsia. 1995; 36:899–904. [PubMed: 7649129]

Lezak, MD. Neuropsychological assessment. 3rd ed.. New York: Oxford University Press; 1995.

Lindenberger U, Potter U. The complex nature of unique and shared effects in hierarchical linear regression: Implications for developmental psychology. Psychological Methods. 1998; 3:218–230.

Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. Am Psychol. 1993; 48:1181–1209. [PubMed: 8297057]

Luria, AR. Higher Cortical Functions in Man. New York: Basic Books; 1966.

MacLullich AM, Ferguson KJ, Deary IJ, Seckl JR, Starr JM, Wardlaw JM. Intracranial capacity and brain volumes are associated with cognition in healthy elderly men. Neurology. 2002; 59:169–174. [PubMed: 12136052]

Madden DJ, Bennett IJ, Song AW. Cerebral white matter integrity and cognitive aging: contributions from diffusion tensor imaging. Neuropsychol Rev. 2009; 19:415–435. [PubMed: 19705281]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol. 2000; 41:49–100. [PubMed: 10945922]

Morgen K, Sammer G, Courtney SM, Wolters T, Melchior H, Blecker CR, Oschmann P, Kaps M, Vaitl D. Evidence for a direct association between cortical atrophy and cognitive impairment in relapsing-remitting MS. Neuroimage. 2006; 30:891–898. [PubMed: 16360321]

Nakamura M, Nestor PG, Levitt JJ, Cohen AS, Kawashima T, Shenton ME, McCarley RW. Orbitofrontal volume deficit in schizophrenia and thought disorder. Brain. 2008; 131:180–195. [PubMed: 18056163]

Nestor PG, Kubicki M, Nakamura M, Niznikiewicz M, McCarley RW, Shenton ME. Comparing prefrontal gray and white matter contributions to intelligence and decision making in schizophrenia and healthy controls. Neuropsychology. 2010; 24:121–129. [PubMed: 20063953]

Orwin RG. A Fail-Safe N for Effect Size in Meta-Analysis Journal of Educational Statistics. 1983; 8:157–159.

Owen AM, McMillan KM, Laird AR, Bullmore E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp. 2005; 25:46–59. [PubMed: 15846822]

Paul R, Grieve SM, Chaudary B, Gordon N, Lawrence J, Cooper N, Clark CR, Kukla M, Mulligan R, Gordon E. Relative contributions of the cerebellar vermis and prefrontal lobe volumes on cognitive function across the adult lifespan. Neurobiol Aging. 2009; 30:457–465. [PubMed: 17869383]

Petrides, M.; Pandya, DN. Association Pathways of the Prefrontal Cortex and Functional Observations. In: Stuss, DT.; Knight, RT., editors. Principles of Frontal Lobe Function. New York: Oxford University Press; 2002. p. 31-50.

Qiu LR, Germann J, Spring S, Alm C, Vousden DA, Palmert MR, Lerch JP. Hippocampal volumes differ across the mouse estrous cycle, can change within 24 hours, and associate with cognitive strategies. Neuroimage. 2013; 83:593–598. [PubMed: 23831531]

Raz N, Briggs SD, Marks W, Acker JD. Age-related deficits in generation and manipulation of mental images: II. The role of dorsolateral prefrontal cortex. Psychol Aging. 1999; 14:436–444. [PubMed: 10509698]

Raz N, Gunning FM, Head D, Dupuis JH, McQuain J, Briggs SD, Loken WJ, Thornton AE, Acker JD. Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. Cereb Cortex. 1997; 7:268–282. [PubMed: 9143446]

Raz N, Gunning-Dixon FM, Head D, Dupuis JH, Acker JD. Neuroanatomical correlates of cognitive aging: evidence from structural magnetic resonance imaging. Neuropsychology. 1998; 12:95–114. [PubMed: 9460738]

Raz, N.; Kennedy, KM. A Systems Approach to Age-Related Change: Neuroanatomic changes, their modifiers, and cognitive correlates. In: Jagust, W.; D'Esposito, M., editors. Imaging the Aging Brain. New York, NY: Oxford University Press; 2009. p. 43-70.

Raz N, Lindenberger U. Only Time Will Tell: Cross-Sectional Studies Offer No Solution to the Age-Brain-Cognition Triangle: Comment on Salthouse (2011). Psychological Bulletin. 2011; 137:790–795. [PubMed: 21859179]

Raz N, Rodrigue KM, Acker JD. Hypertension and the brain: vulnerability of the prefrontal regions and executive functions. Behav Neurosci. 2003; 117:1169–1180. [PubMed: 14674838]

Rhodes MG. Age-related differences in performance on the Wisconsin card sorting test: a meta-analytic review. Psychol Aging. 2004; 19:482–494. [PubMed: 15382998]

Rodriguez-Aranda C, Martinussen M. Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Word Association Task (COWAT): a meta-analytic study. Dev Neuropsychol. 2006; 30:697–717. [PubMed: 16995832]

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB. Modelling neural correlates of working memory: a coordinate-based meta-analysis. Neuroimage. 2012; 60:830–846. [PubMed: 22178808]

Rusch N, Tebartz van Elst L, Valerius G, Buchert M, Thiel T, Ebert D, Hennig J, Olbrich HM. Neurochemical and structural correlates of executive dysfunction in schizophrenia. Schizophr Res. 2008; 99:155–163. [PubMed: 17616347]

Salat DH, Kaye JA, Janowsky JS. Greater orbital prefrontal volume selectively predicts worse working memory performance in older adults. Cereb Cortex. 2002; 12:494–505. [PubMed: 11950767]

Salthouse TA. Decomposing adult age differences in working memory. Developmental psychology. 1991; 27:763–776.

Saxby BK, Harrington F, McKeith IG, Wesnes K, Ford GA. Effects of hypertension on attention, memory, and executive function in older adults. Health Psychol. 2003; 22:587–591. [PubMed: 14640855]

Schretlen D, Pearlson GD, Anthony JC, Aylward EH, Augustine AM, Davis A, Barta P. Elucidating the contributions of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. J Int Neuropsychol Soc. 2000; 6:52–61. [PubMed: 10761367]

Spreen, O.; Strauss, E. A compendium of neuropsychological tests. 2nd ed.. New York: Oxford University Press; 1998.

Strick PL, Dum RP, Fiez JA. Cerebellum and nonmotor function. Annu Rev Neurosci. 2009; 32:413–434. [PubMed: 19555291]

Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology. 1935; 18:643–662.

Stuss DT, Alexander MP. Executive functions and the frontal lobes: a conceptual view. Psychol Res. 2000; 63:289–298. [PubMed: 11004882]

Stuss, DT.; Alexander, MP.; Floden, D.; Binns, MA.; Levine, B.; McIntosh, AR.; Rajah, N.; Hevenor, SJ. Fractionation and Localization of Distinct Frontal Lobe Processes: Evidence from Focal Lesions in Humans. In: Stuss, DT.; Knight, RT., editors. Principles of Frontal Lobe Function. New York: Oxford University Press; 2002. p. 392-407.

Stuss DT, Shallice T, Alexander MP, Picton TW. A Multidisciplinary Approach to Anterior Attentional Functionsa. Annals of the New York Academy of Sciences. 1995; 769:191–212. [PubMed: 8595026]

Takeuchi H, Taki Y, Sassa Y, Hashizume H, Sekiguchi A, Fukushima A, Kawashima R. Brain structures associated with executive functions during everyday events in a nonclinical sample. Brain Struct Funct. 2013; 218:1017–1032. [PubMed: 22851058]

Teuber HL. Unity and diversity of frontal lobe functions. Acta Neurobiol Exp (Wars). 1972; 32:615–656. [PubMed: 4627626]

Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol. 2004; 19:203–214. [PubMed: 15010086]

Tu PC, Chen LF, Hsieh JC, Bai YM, Li CT, Su TP. Regional cortical thinning in patients with major depressive disorder: a surface-based morphometry study. Psychiatry Res. 2012; 202:206–213. [PubMed: 22521631]

Tullberg M, Fletcher E, DeCarli C, Mungas D, Reed BR, Harvey DJ, Weiner MW, Chui HC, Jagust WJ. White matter lesions impair frontal lobe function regardless of their location. Neurology. 2004; 63:246–253. [PubMed: 15277616]

Van Petten C, Plante E, Davidson PS, Kuo TY, Bajuscak L, Glisky EL. Memory and executive function in older adults: relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. Neuropsychologia. 2004; 42:1313–1335. [PubMed: 15193940]

Vicario A, Martinez CD, Baretto D, Diaz Casale A, Nicolosi L. Hypertension and cognitive decline: impact on executive function. J Clin Hypertens (Greenwich). 2005; 7:598–604. [PubMed: 16227762]

Weigl E. On the psychology of so-called processes of abstraction. Journal of abnormal and social psychology. 1941; 36:3–33.

Yendiki A, Greve DN, Wallace S, Vangel M, Bockholt J, Mueller BA, Magnotta V, Andreasen N, Manoach DS, Gollub RL. Multi-site characterization of an fMRI working memory paradigm: reliability of activation indices. Neuroimage. 2010; 53:119–131. [PubMed: 20451631]

Zimmerman ME, Brickman AM, Paul RH, Grieve SM, Tate DF, Gunstad J, Cohen RA, Aloia MS, Williams LM, Clark CR, Whitford TJ, Gordon E. The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan. Am J Geriatr Psychiatry. 2006; 14:823–833. [PubMed: 17001022]

Zuffante P, Leonard CM, Kuldau JM, Bauer RM, Doty EG, Bilder RM. Working memory deficits in schizophrenia are not necessarily specific or associated with MRI-based estimates of area 46 volumes. Psychiatry Res. 2001; 108:187–209. [PubMed: 11756016]

Highlights

We examined strength of association of executive functions & PFC size in 28 studies.

Larger PFC volume and greater thickness correlated with better executive functions.

Stronger correlations found in samples with wider age dispersion.

Effect size depended on executive task type and methods of volume estimation.

Effect Sizes for PFC volume

Author and Year		Effect Size (Zr) [95% CI]

Salat, 2002	−0.25 [−0.59 , 0.09]
Takeuchi, 2013	−0.19 [−0.36 , −0.03]
Takeuchi, 2013	−0.15 [−0.30 , 0.00]
Zuffante, 2001	−0.11 [−0.52 , 0.30]
Gianaros, 2006	−0.04 [−0.30 , 0.22]
Baare, 1999	0.00 [−0.54 , 0.54]
Garlinghouse, 2010	0.00 [−0.39 , 0.39]
Morgen, 2006	0.00 [−0.46 , 0.46]
Van Petten, 2004	0.00 [−0.29 , 0.29]
Choi, 2004	0.00 [−0.34 , 0.34]
Koutsouleris, 2010	0.00 [−0.36 , 0.36]
Kozicky, 2013	0.00 [−0.36 , 0.36]
Paul, 2009	0.04 [−0.08 , 0.16]
Hanninen, 1997	0.07 [−0.22 , 0.36]
Gianaros, 2006	0.09 [−0.13 , 0.32]
Gautam, 2011	0.12 [0.02 , 0.22]
Gautam, 2011	0.12 [0.02 , 0.22]
MacLullich, 2002	0.13 [−0.07 , 0.33]
Nakamura, 2008	0.18 [−0.25 , 0.60]
Schretlen, 2000	0.21 [0.04 , 0.39]
Tullberg, 2004	0.23 [−0.03 , 0.49]
Rusch, 2008	0.23 [−0.14 , 0.60]
Gur, 2000	0.28 [−0.03 , 0.60]
Gunning−Dixon, 2003	0.33 [0.18 , 0.48]
Zimmerman, 2006	0.33 [0.18 , 0.47]
Raz, 1998	0.37 [0.19 , 0.54]
Head, 2009	0.41 [0.25 , 0.56]
Haldane, 2008	0.41 [0.16 , 0.66]
Goldstein, 2011	0.44 [0.04 , 0.85]
Gur, 2000	0.45 [0.22 , 0.68]
Bender, 2012	0.56 [0.40 , 0.72]
RE Model	0.15 [0.08 , 0.23]

Figure 1.
Effect sizes (correlations) for associations between executive functions and PFC volume.

Effect Sizes for Cortical thickness

Author and Year	Effect Size (Zr) [95% CI]

Gautam, 2011 — 0.00 [−0.10 , 0.10]
Dickerson, 2008 — 0.00 [−0.52 , 0.52]
Grambaite, 2011 — 0.00 [−0.42 , 0.42]
Kochunov, 2009 — 0.00 [−0.34 , 0.35]
Hartberg, 2011 — 0.01 [−0.13 , 0.15]
Gautam, 2011 — 0.05 [−0.05 , 0.15]
Burzynska, 2012 — 0.12 [−0.10 , 0.35]
Tu, 2012 — 0.25 [−0.07 , 0.56]
Kochunov, 2009 — 0.35 [0.07 , 0.63]
Burzynska, 2012 — 0.37 [0.14 , 0.60]

RE Model — 0.10 [0.01 , 0.18]

−1   −0.5   0   0.5   1
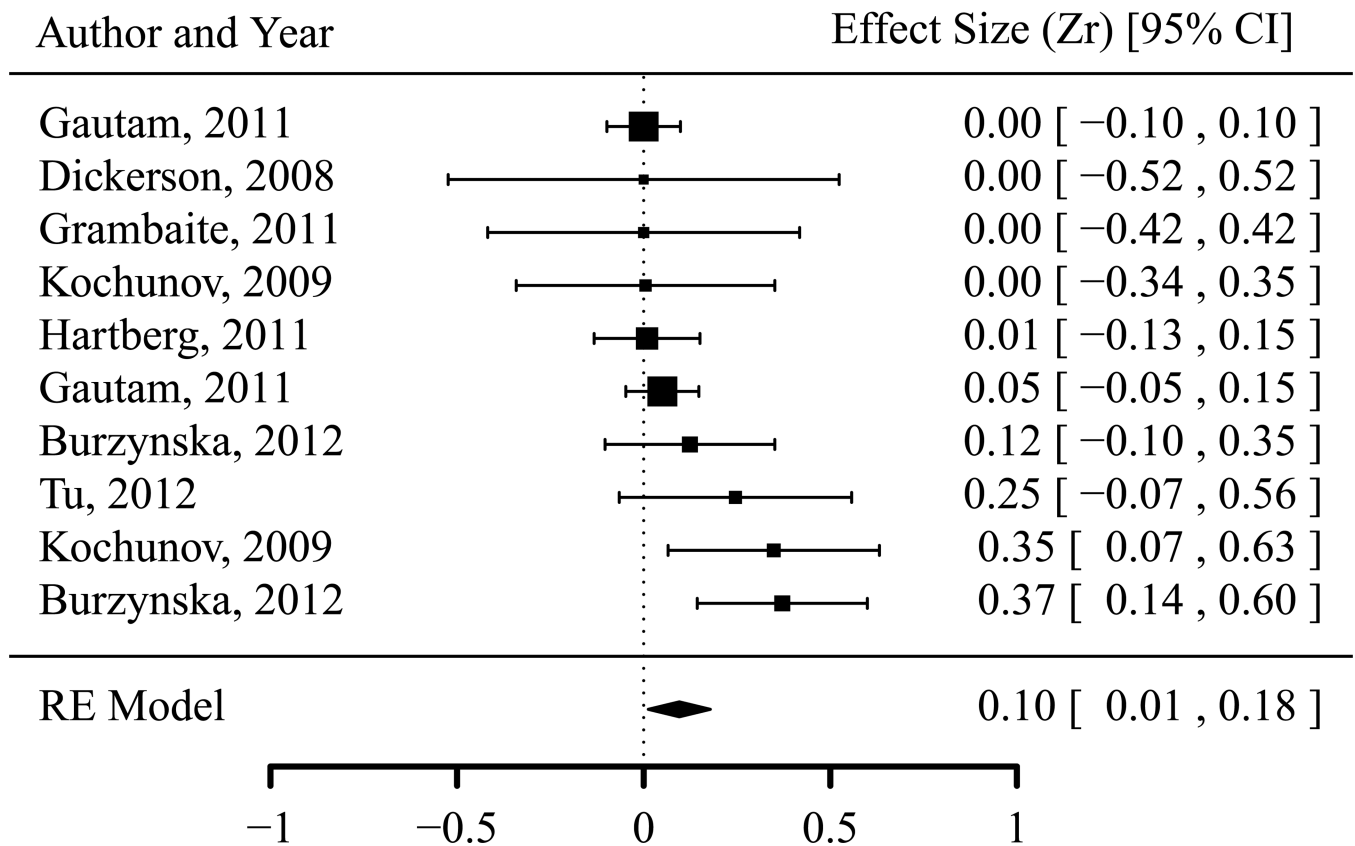
Figure 2.
Effect sizes (correlations) for associations between executive functions and prefrontal cortical thickness.
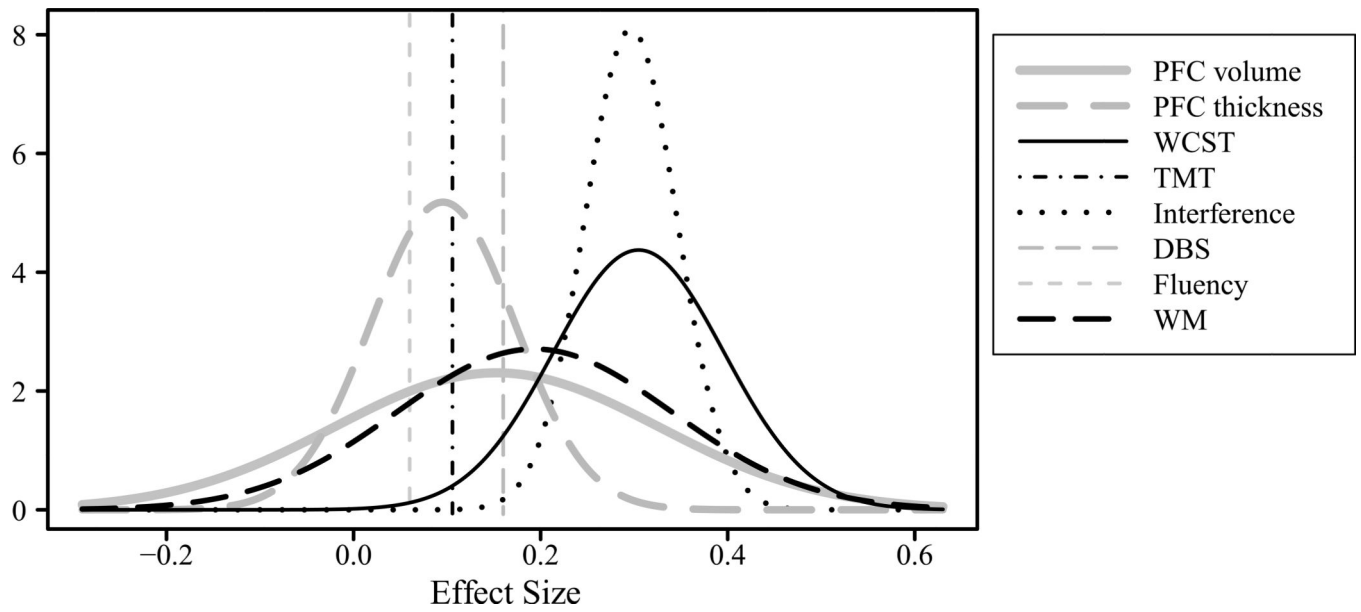
Figure 3.
Estimated distributions of true effect sizes.
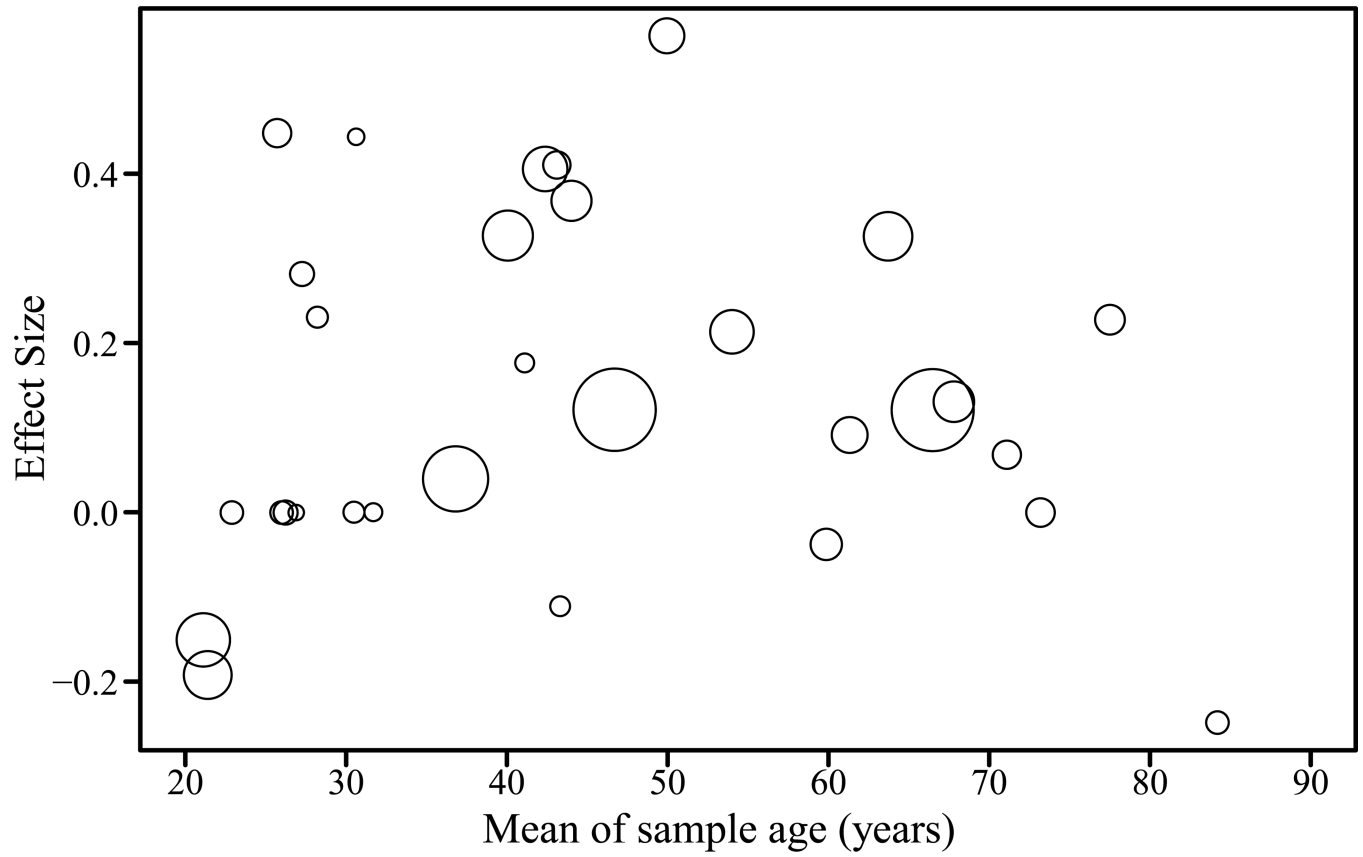
Figure 4.
Scatterplot of effect size vs. sample mean age. The areas of circles are proportional to sample sizes.
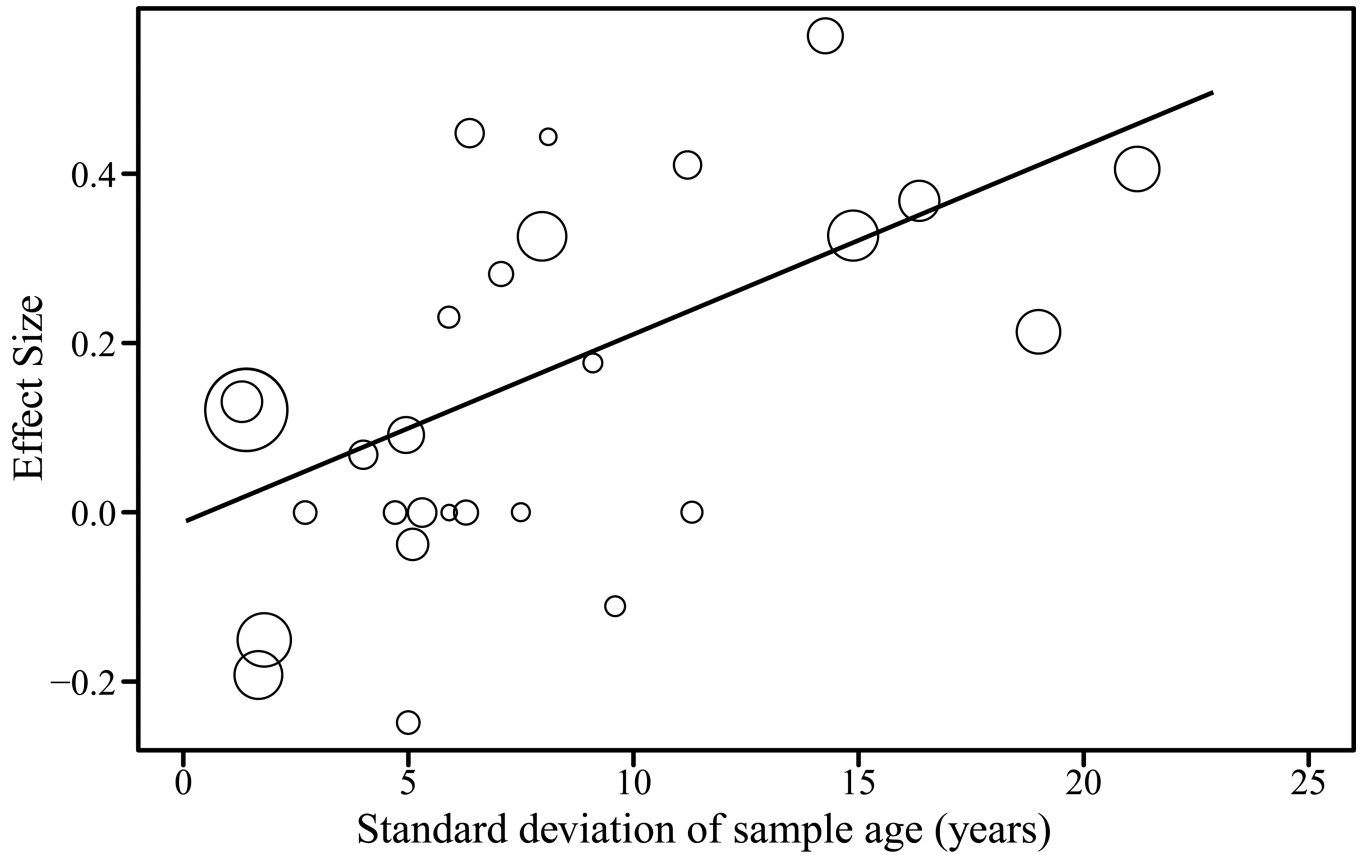
Figure 5.
Standard deviation of sample age as a moderator. The areas of circles are proportional to sample sizes.

Effect Sizes: WCST

Author and Year		Effect Size (Zr) [95% CI]

Nakamura, 2008   0.00 [−0.44 , 0.44]

Choi, 2004   0.00 [−0.34 , 0.34]

Hanninen, 1997   0.14 [−0.14 , 0.43]

Schretlen, 2000   0.21 [0.04 , 0.39]

Rusch, 2008   0.23 [−0.14 , 0.60]

Gur, 2000   0.28 [−0.03 , 0.60]

Gunning−Dixon, 2003   0.31 [0.16 , 0.46]

Raz, 1998   0.45 [0.29 , 0.61]

Gur, 2000   0.45 [0.22 , 0.68]

Head, 2009   0.45 [0.30 , 0.59]

RE Model   0.30 [0.21 , 0.40]

−1   −0.5   0   0.5   1
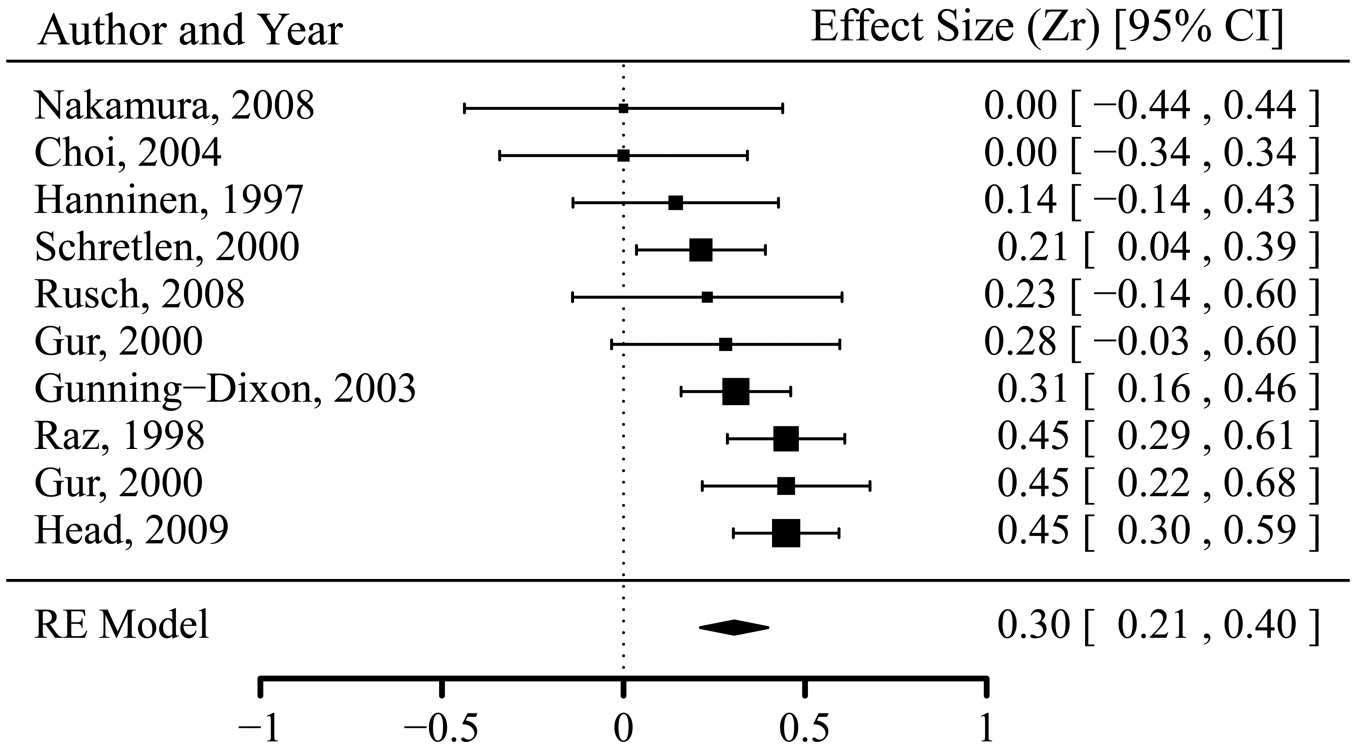
Figure 6.
Effect sizes (correlations) for associations between WCST performance and PFC volume.

Effect Sizes: Working Memory

Author and Year		Effect Size (Zr) [95% CI]
Salat, 2002		−0.25 [−0.59 , 0.09]
Zuffante, 2001		−0.11 [−0.52 , 0.30]
Gianaros, 2006		−0.08 [−0.34 , 0.18]
Garlinghouse, 2010		0.00 [−0.39 , 0.39]
Morgen, 2006		0.00 [−0.46 , 0.46]
Kozicky, 2013		0.00 [−0.36 , 0.36]
Gianaros, 2006		0.10 [−0.13 , 0.32]
Gautam, 2011		0.15 [0.05 , 0.24]
Gautam, 2011		0.16 [0.06 , 0.26]
Paul, 2009		0.16 [0.04 , 0.28]
Zimmerman, 2006		0.23 [0.08 , 0.39]
Raz, 1998		0.29 [0.10 , 0.47]
Gunning−Dixon, 2003		0.34 [0.20 , 0.49]
Goldstein, 2011		0.44 [0.04 , 0.85]
Head, 2009		0.46 [0.31 , 0.60]
Bender, 2012		0.56 [0.40 , 0.72]
RE Model		0.19 [0.10 , 0.29]

−1   −0.5   0   0.5   1

Figure 7.
Effect sizes (correlations) for associations between WM and PFC volume.

Effect Sizes: Interference test
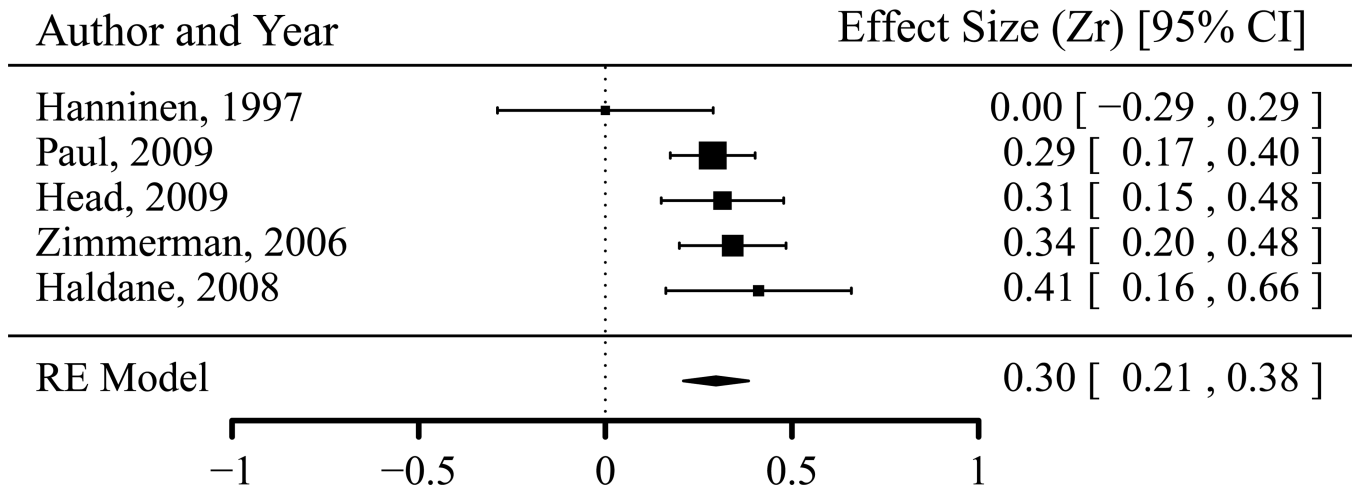
Figure 8.
Effect sizes (correlations) for associations between interference score and PFC volume.

Effect Sizes: Fluency test

Author and Year	Effect Size (Zr) [95% CI]

Baare, 1999   0.00 [−0.54 , 0.54]

Choi, 2004   0.00 [−0.34 , 0.34]

Paul, 2009   0.03 [−0.09 , 0.15]

Hanninen, 1997   0.13 [−0.16 , 0.41]

MacLullich, 2002   0.13 [−0.07 , 0.33]

RE Model   0.06 [−0.03 , 0.15]
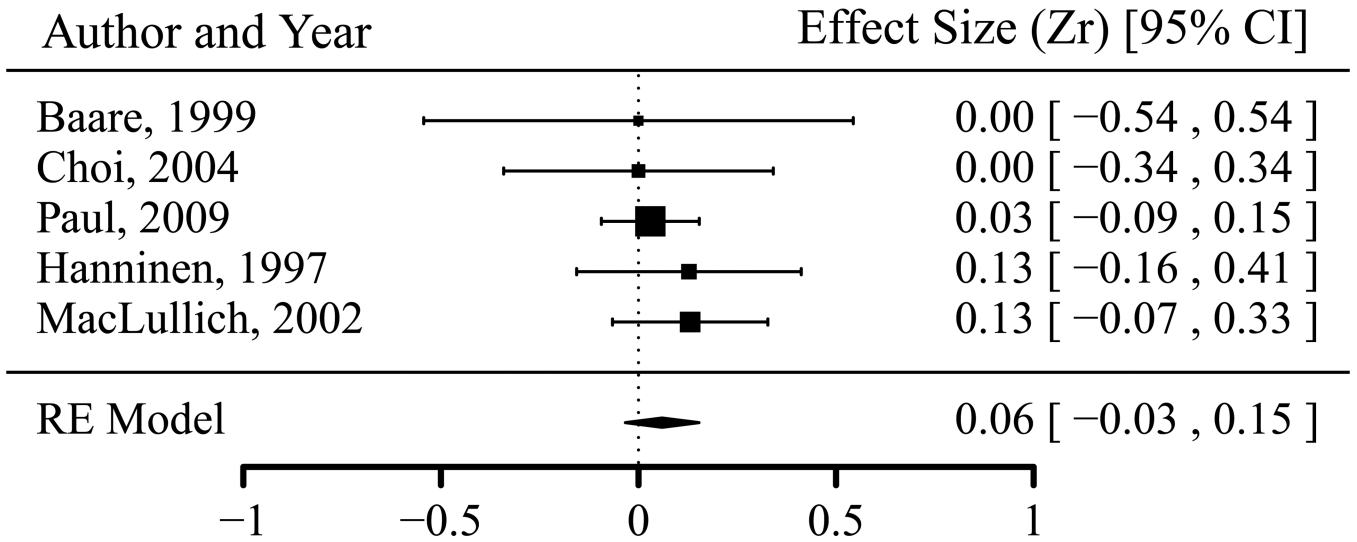
Figure 9.
Effect sizes (correlations) for associations between fluency test score and PFC volume.

Effect Sizes: TMT

Author and Year		Effect Size (Zr) [95% CI]
Paul, 2009		−0.32 [−0.43 , −0.21]
Choi, 2004		0.00 [−0.34 , 0.34]
Hanninen, 1997		0.00 [−0.29 , 0.29]
Koutsouleris, 2010		0.00 [−0.36 , 0.36]
Gianaros, 2006		0.00 [−0.25 , 0.26]
Gautam, 2011		0.08 [−0.02 , 0.18]
Gianaros, 2006		0.09 [−0.14 , 0.31]
Gautam, 2011		0.09 [0.00 , 0.19]
Zimmerman, 2006		0.26 [0.11 , 0.41]
Nakamura, 2008		0.35 [−0.03 , 0.74]
RE Model		0.05 [−0.09 , 0.18]

−1   −0.5   0   0.5   1
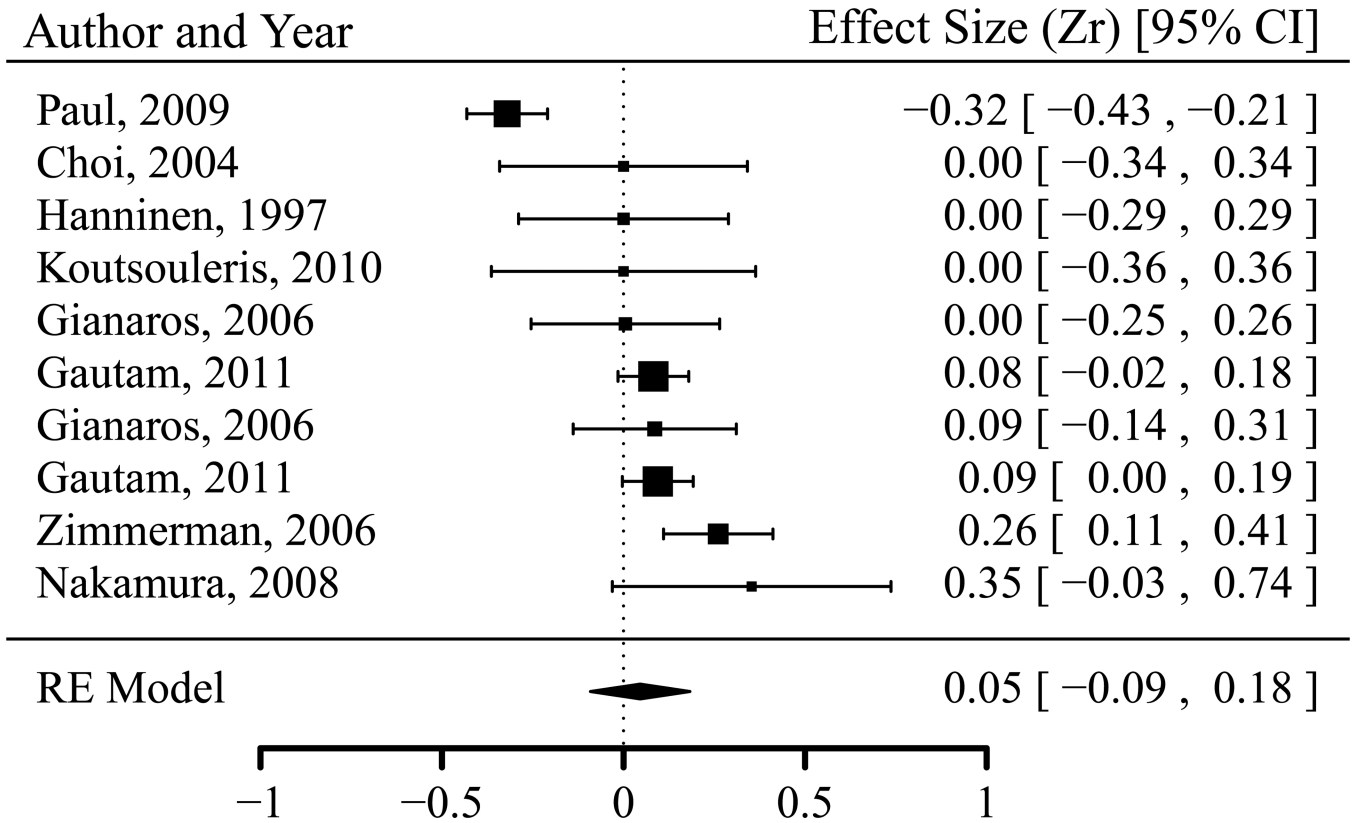
Figure 10.
Effect sizes (correlations) for associations between TMT score and PFC volume.

Effect Sizes: LPFC

Author and Year		Effect Size (Zr) [95% CI]
Zuffante, 2001		−0.11 [−0.52 , 0.30]
Gianaros, 2006		−0.04 [−0.30 , 0.22]
Baare, 1999		0.00 [−0.54 , 0.54]
Morgen, 2006		0.00 [−0.46 , 0.46]
Van Petten, 2004		0.00 [−0.29 , 0.29]
Salat, 2002		0.00 [−0.36 , 0.36]
Kozicky, 2013		0.00 [−0.36 , 0.36]
Paul, 2009		0.04 [−0.08 , 0.16]
Gianaros, 2006		0.09 [−0.13 , 0.32]
Gautam, 2011		0.12 [0.02 , 0.22]
Gautam, 2011		0.12 [0.02 , 0.22]
Rusch, 2008		0.23 [−0.14 , 0.60]
Zimmerman, 2006		0.34 [0.20 , 0.48]
Head, 2009		0.41 [0.25 , 0.56]
Haldane, 2008		0.41 [0.16 , 0.66]
Goldstein, 2011		0.44 [0.04 , 0.85]
Gur, 2000		0.47 [0.25 , 0.70]
Gur, 2000		0.56 [0.33 , 0.80]
Bender, 2012		0.56 [0.40 , 0.72]
RE Model		0.22 [0.12 , 0.31]

−1   −0.5   0   0.5   1

Figure 11.
Effect sizes (correlations) for associations between executive functions and LPFC volume.

Effect Sizes: OFC

Author and Year	Effect Size (Zr) [95% CI]

Salat, 2002	−0.50 [−0.77 , −0.22]
Takeuchi, 2013	−0.19 [−0.36 , −0.03]
Takeuchi, 2013	−0.15 [−0.30 , 0.00]
Baare, 1999	0.00 [−0.54 , 0.54]
Morgen, 2006	0.00 [−0.46 , 0.46]
Choi, 2004	0.00 [−0.34 , 0.34]
Nakamura, 2008	0.18 [−0.25 , 0.60]
Zimmerman, 2006	0.29 [0.15 , 0.44]
RE Model	−0.06 [−0.26 , 0.15]
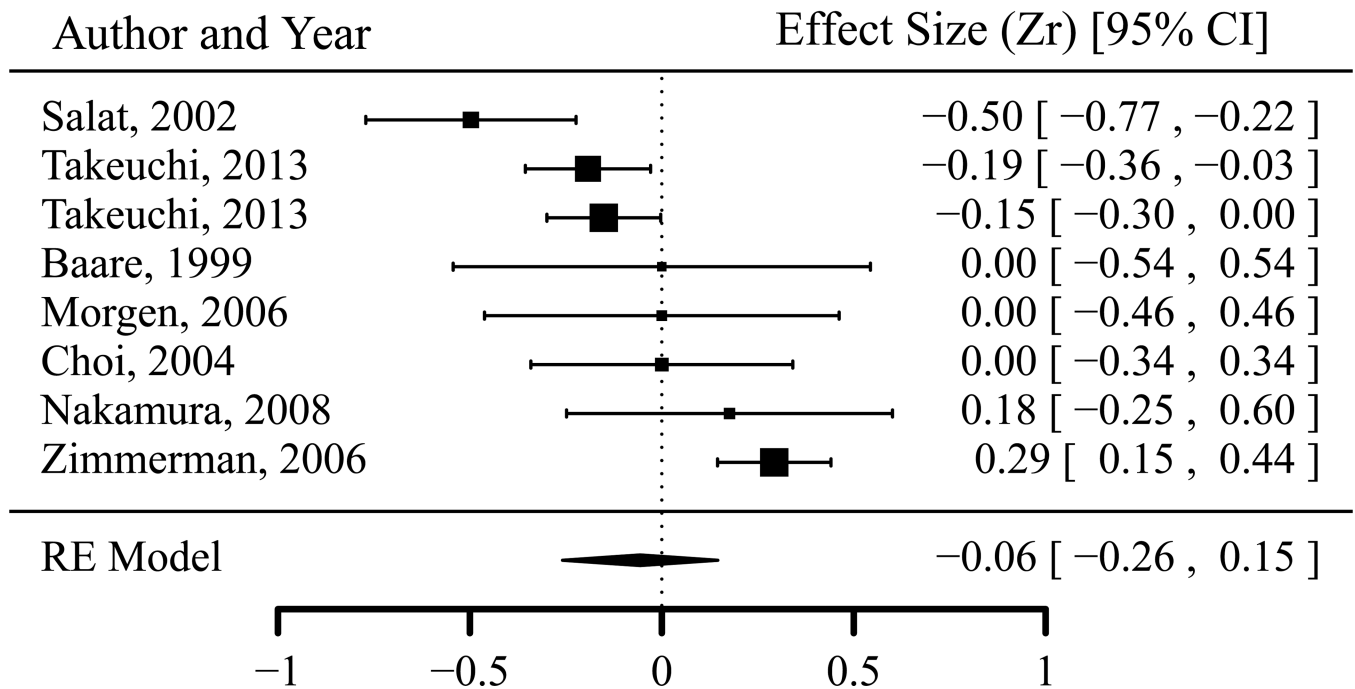
Figure 12.
Effect sizes (correlations) for associations between executive functions and OFC volume.

Effect Sizes: MPFC

Author and Year	Effect Size (Zr) [95% CI]
Morgen, 2006	0.00 [−0.46 , 0.46]
Gur, 2000	0.00 [−0.34 , 0.34]
Koutsouleris, 2010	0.00 [−0.36 , 0.36]
Baare, 1999	0.00 [−0.54 , 0.54]
Zimmerman, 2006	0.35 [0.21 , 0.49]
Gur, 2000	0.42 [0.19 , 0.66]
RE Model	0.20 [0.03 , 0.37]

−1   −0.5   0   0.5   1

Figure 13.
Effect sizes (correlations) for associations between executive functions and MPFC volume.

Table 1

Study Characteristics.

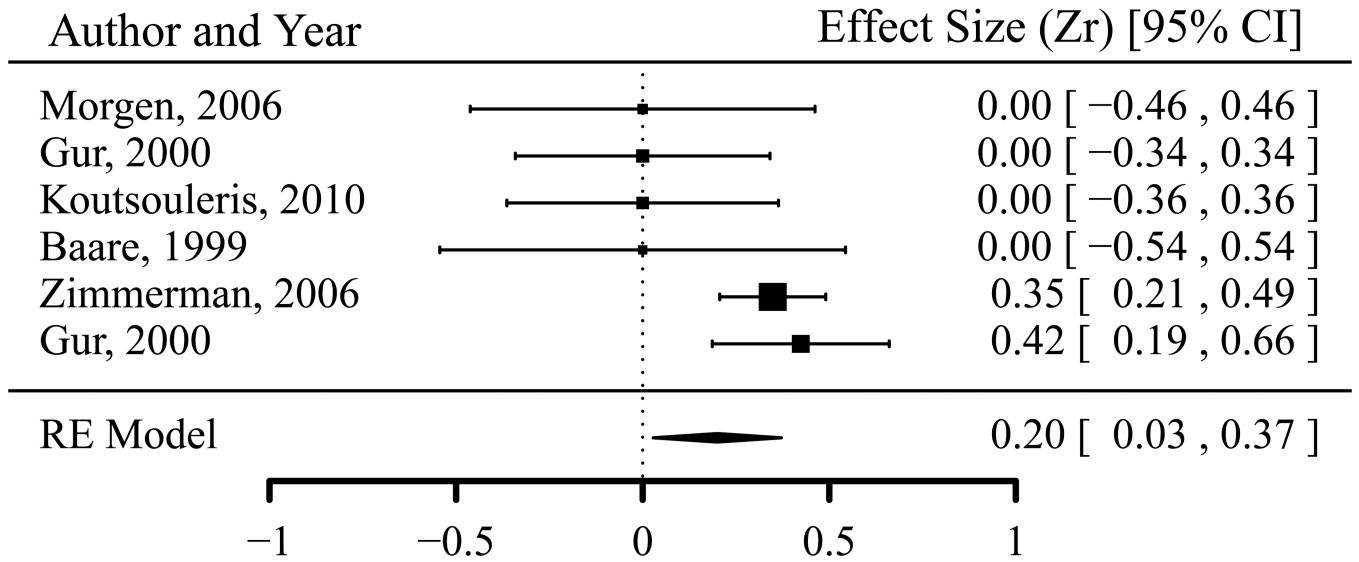| Study	Sample Size	Age[a]	% female	Structure measure[b]	Method	Regions[c]	Cognitive Test[d]
Baare, 1999	14	26.9 (5.9)	0	lobe volume	ROI	PFC, LPFC, OFC, MPFC	verbal & semantic fluency
Bender, 2012	72	49.96 (14.26) [19–77]	69.4%	GM volume	ROI	LPFC	Composite WM score
Burzynska, 2012	56	64.8 (2.6) [60–71]	48.2%	GM thickness	VBM	LPFC	WCST
	73	25.6 (3.1) [20–32]	43.8%				
Choi, 2004	34	26.24 (6.28)	17.7%	GM volume	ROI	OFC	TMT; WCST; verbal fluency
Dickerson, 2008	15	69.5 (4.8) [66–81]	46.7%	GM thickness	VBM	PFC	TMT
Garlinghouse, 2010	26	30.5 (11.3)	50.0%	lobe volume	ROI	frontal lobe	DBS (WM)
Gautam, 2011	397	66.5 (1.40) [64–68]	44.6%	GM volume GM thickness	ROI	LPFC	TMT; DBS (WM)
	400	46.7 (1.40) [44–48]	55.5%				
Gianaros, 2006	58	59.86 (5.10) [52–70]	100%	GM volume	ROI	LPFC	TMT; Fourword Short-term Memory Test (WM)
	76	61.33 (4.95) [50–70]	0				
Goldstein, 2011	16	30.63 (8.11)	43.8%	GM volume	ROI	LPFC	CANTAB (WM)
Grambaite, 2011	23	63.2 (8.0)	52.2%	GM thickness	ROI	LPFC, OFC	TMT; verbal fluency; Interference
Gunning-Dixon, 2003	139	63.71 (7.97) [50–81]	59.7%	GM volume	ROI	DLPFC + OFC	WCST; computation span and listening span (WM)
Gur, 2000	47	25.72 (6.36)	100%	GM volume	ROI	LPFC & MPFC	WCST
	34	27.26 (7.07)	0				
Haldane, 2008	44	43.1 (11.2)	54.5%	GM volume	VBM	LPFC	Interference: Stroop
Hanninen, 1997	47	71.1 (4.0)	51.1%	lobe volume	ROI	frontal lobe	verbal fluency; Interference: Stroop; WCST; TMT
Hartberg, 2011	192	36.1 (9.6)	49.0%	GM thickness	ROI	LPFC	digit span (WM), color-word interference, set shifting
Head, 2009	117	42.4 (21.2) [18–80]	70.94%	GM volume	ROI	DLPFC	WCST; WM; Interference
Kochunov, 2009	38	[30–90]	60.5%	GM thickness	ROI	LPFC	Executive Interview
	33	[19–29]	63.6%				
Koutsouleris, 2010	30	26.0 (2.7)	40.0%	GM volume	VBM	MPFC	TMT
Kozicky, 2013	30	22.9 (4.7)	60%	GM volume	ROI	DLPFC	CANTAB (WM)
MacLullich, 2002	97	67.8 (1.3) [65–70]	--	lobe volume	ROI	frontal lobe	verbal fluency
Morgen, 2006	19	31.7 (7.5) [22–44]	--	GM volume	VBM	PFC	DBS (WM)
Nakamura[e], 2008	21	41.1 (9.1) [22–55]	23.8%	GM volume	ROI	OFC	TMT; WCST

Study	Sample Size	Age[a]	% female	Structure measure[b]	Method	Regions[c]	Cognitive Test[d]
Paul, 2009	251	36.8 [18–79]	58.2%	GM volume	ROI	LPFC	TMT; Interference; Fluency; DBS (WM)
Raz, 1998	95	44.02 (16.35) [18–77]	56.80%	GM volume	ROI	DLPFC+OFC	WCST; composite WM
Rusch, 2008	26	28.2 (5.9)	32.3%	GM volume	ROI	DLPFC	WCST
Salat, 2002	30	84.2 [71.6–93.5]	53.3%	GM volume	ROI	LPFC, OFC	self-ordered pointing and n-back (WM)
Schretlen, 2000	112	54 (19)	57.1%	lobe volume	ROI	frontal lobe	modified WCST
Takeuchi, 2013	168	21.12 (1.79)	0.00%	GM volume	VBM	OFC	Dysexecutive Questionnaire
	135	21.39 (1.67)	100.00%				
Tu, 2012	36	41.81 (10.70)	61.1%	GM thickness	VBM	LPFC	Color Trail Test (substitute TMT)
Tullberg, 2004	52	77.5	--	GM volume	ROI	PFC	Composite executive function: Perseveration, DBS, visual span backwards, fluency test.
Van Petten, 2004	48	73.2 (5.3) [65.5–85.8]	68.8%	GM volume	ROI	LPFC	F-factor, including WCST, verbal fluency, mental arithmetic, DBS and mental control tests
Zimmerman, 2006	148	40.05 (14.88)	49%	GM volume	ROI	LPFC, MPFC, OFC	TMT-B; Stroop interference; DBS (WM); maze
Zuffante, 2001	23	43.3 (9.6)	0	lobe volume	ROI	LPFC	spatial delayed response task, self-ordered pointing (WM)

Note:

[a] Mean ages are shown, with standard deviation of age in parentheses and age range in brackets.

[b] GM = gray matter; lobe volume = volume of gray matter + white matter;

[c] LPFC = lateral prefrontal cortex; OFC = orbito-frontal cortex; MPFC = medial prefrontal cortex.

[d] WCST = Wisconsin Card Sorting Test; WM = working memory; TMT = Trail Making Test; DBS = Digit backwards span; CANTAB: Cambridge Neuropsychological Test Automated Battery.

[e] In Nakamura's (2008) study, age was averaged from 25 subjects, whereas the cognition-volume relationship was assessed from a subset of 21 subjects.

Table 2

Summary effect sizes within each age group.

Age group (mean age)	k	N	Weighted Effect Size			95% CI for M*		mean weighted correlation (r)	Homogeneity Tests		
			mean (M*)	se	p				Q_w	df	p
Overall	31	2806	0.153	0.039	<.001	0.077	0.230	0.152	118.03	30	<.001
20–29	9	518	0.060	0.088	0.495	−0.112	0.232	0.060	28.31	8	<.001
30–39	4	312	0.082	0.079	0.301	−0.073	0.237	0.082	3.68	3	0.299
40–50	8	920	0.314	0.070	<.001	0.178	0.450	0.304	31.04	7	<.001
50–59	2	170	0.106	0.124	0.393	−0.137	0.350	0.106	2.45	1	0.117
60–69	4	709	0.173	0.056	0.002	0.063	0.282	0.171	5.82	3	0.121
70–79	3	147	0.107	0.082	0.190	−0.053	0.267	0.107	1.43	2	0.489
>80	1	30	−0.249	0.174	0.153	−0.590	0.093	−0.244	--	--	--

Note: Positive effect sizes refer to the relationship that larger prefrontal volume is associated with better executive function. k = number of effect sizes included in the analysis; N = cumulative sample size across studies; se = standard error of weighted mean; CI = confidence interval.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Summary effect sizes within each domain of cognitive tests.

| Cognitive Test | k | N | Weighted Effect Size | | | 95% CI for M* | | mean weighted correlation (r) | Homogeneity Test | | |
			mean (M*)	se	p				Q_w	df	p
WCST	10	672	0.305	0.048	<.001	0.212	0.398	0.296	15.092	9	0.088
Fluency	5	443	0.060	0.048	0.207	−0.033	0.153	0.060	1.107	1	0.893
Interference	5	607	0.297	0.045	<.001	0.209	0.384	0.288	5.292	4	0.259
TMT (all samples)	10	1462	0.046	0.070	0.511	−0.091	0.183	0.046	52.886	9	<.001
TMT (exclude outlier)	9	1211	0.106	0.028	<.001	0.050	0.161	0.106	7.758	8	0.458
Working Memory	16	1897	0.193	0.047	<.001	0.101	0.286	0.191	56.602	15	<.001
Digit backwards span	6	1241	0.160	0.028	<.001	0.106	0.214	0.159	2.029	5	0.845

Note: Positive effect sizes refer to the relationship that larger prefrontal volume is associated with better executive function. k = number of effect sizes included in the analysis; N = cumulative sample size across studies; se = standard error of weighted mean; CI = confidence interval.

Table 4

Summary effect sizes within each sub-region of prefrontal area.

Sub-region	k	N	Weighted Effect Size			95% CI for M*		mean weighted correlation (r)	Homogeneity Test		
			mean (M*)	se	p				Q_w	df	p
LPFC	19	1850	0.217	0.050	<.001	0.121	0.314	0.214	72.941	18	<.001
OFC	8	569	−0.056	0.103	0.585	−0.259	0.146	−0.056	36.820	7	<.001
MPFC	6	292	0.200	0.088	0.023	0.028	0.372	0.197	9.565	5	0.089

Note: Positive effect sizes refer to the relationship that larger prefrontal volume is associated with better executive function. k = number of effect sizes included in the analysis; N = cumulative sample size across studies; se = standard error of weighted mean; CI = confidence interval.

Table 5

Correlation with executive functions in regional non-PFC volumes: numbers of studies finding positive and null associations.

	WCST		TMT		Interference		WM		Fluency		others	
	P	**0**	**P**	**0**	**P**	**0**	**P**	**0**	**P**	**0**	**P**	**0**
Parietal	—	1	—	—	—	—	—	2	—	—	—	—
Temporal	1	—	—	1	—	—	1	2	—	1	—	—
Occipital	1	—	—	—	—	—	1	—	—	—	—	1
Fusiform Gyrus	1	—	—	—	—	—	1	—	—	—	—	—
Cingulate cortex	—	—	—	1	—	—	—	2	—	—	—	—
Hippocampus	—	1	—	—	—	—	1	—	—	1	—	—
Cerebellum	—	—	1	—	—	1	1	—	—	1	—	—

Note: P: positive relationship; 0: null relationship; —: N/A, relationship not studied in review.