Grey and white matter metrics demonstrate distinct and complementary prediction of differences in cognitive performance in children: Findings from ABCD (N= 11 876)

Lea C. Michel[1], Ethan M. McCormick[1,2,3] & Rogier A. Kievit[1]

1. Cognitive Neuroscience Department, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands
2. Methodology and Statistics, Institute of Psychology, Leiden University, Leiden, Netherlands
3. Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, United States

Corresponding author: Correspondence should be addressed to Lea Michel, Cognitive Neuroscience Department, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands. Emails: lea.michel@radboudumc.nl.

- Number of pages: 36
- Number of figures (9), tables (1), multimedia (0), and 3D models (0)
- Number of words for abstract (249), introduction (650), and discussion (1500)

Conflict of interest: The authors declare no competing financial interests.

Acknowledgements: RAK and LM were supported by a Hypatia Fellowship at the RadboudUMC. We would like to thank Dr Sam Parsons and Dr Nick Judd for their valuable inputs on earlier drafts of the manuscript.

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

Abstract :

Individual differences in cognitive performance in childhood are a key predictor of significant life outcomes such as educational attainment and mental health. Differences in cognitive ability are governed in part by variations in brain structure. However, studies commonly focus on either grey or white matter metrics in humans, leaving open the key question as to whether grey or white matter microstructure play distinct or complementary roles supporting cognitive performance.

To compare the role of grey and white matter in supporting cognitive performance, we used regularized structural equation models to predict cognitive performance with grey and white matter measures. Specifically, we compared how grey matter (volume, cortical thickness and surface area) and white matter measures (volume, fractional anisotropy and mean diffusivity) predicted individual differences in cognitive performance. The models were tested in 11,876 children (ABCD Study, 5680 female; 6196 male) at 10 years old.

We found that grey and white matter metrics bring partly non-overlapping information to predict cognitive performance. The models with only grey or white matter explained respectively 15.4% and 12.4% of the variance in cognitive performance, while the combined model explained 19.0%. Zooming in we additionally found that different metrics within grey and white matter had different predictive power, and that the tracts/regions that were most predictive of cognitive performance differed across metric.

These results show that studies focusing on a single metric in either grey or white matter to study the link between brain structure and cognitive performance are missing a key part of the equation.

Significance Statement:

This paper enriches the recent debates on the challenges of linking variation in brain structure to phenotypic differences (Marek et al., 2022). We demonstrate that using latent variables (to improve power), structural equation modelling (to allow greater flexibility in linking brain to behaviour), and by simultaneously incorporating multiple measures of grey and white matter in a large sample, we demonstrate relatively strong and robust brain-behaviour associations, which highlight the complementarity of grey and white matter metrics in predicting cognitive performance as well as the importance of incorporating the full complexity of these associations over 1-to-1 linkages. This finding should lead researchers to consider integrating both grey and white matter measures when demonstrating a more comprehensive picture of brain-cognition relationships.

Introduction

The field of cognitive neuroscience is premised on the hypothesis that differences in cognitive performance can be understood in part by studying differences in brain structure and function (Basten et al., 2015). Although much is known about the relation between grey and white matter structure and cognitive performance (Tamnes, Østby, Walhovd, et al., 2010; Tamnes, Østby, Fjell, et al., 2010; Magistro et al., 2015; Muetzel et al., 2015; Schnack et al., 2015; Deoni et al., 2016), less is known about how the structure of the two tissues *together* explain differences in cognitive performance.

One way to look at this challenge is to consider grey and white matter as two sides of the same coin. At the cellular level, they are both composed of neurons, although different parts (cell bodies, dendrites, synapses and axons for grey matter and (un)myelinated axons for white matter), glial cells and vasculature (Wandell, 2016; Purves et al., 2019). Moreover, recent findings observe even more overlap, namely that differences in one tissue (white matter myelination) may affect the other tissue (cortical thickness) (Natu et al., 2019), suggesting they may capture the same neurobiological properties using different techniques.

A counter argument would be that grey and white matter are two distinct brain properties with distinct mechanistic roles. Twin study reports that grey and white matter volume shared 68% heritability, suggesting both overlapping and distinct genetic mechanisms (Baaré et al., 2001) and recent attention has focused on their different transcriptome patterns, highlighting the differentiation of their cells and their specific functional roles (Mills et al., 2013). Studies have shown the importance of both dendritic network and myelinated axons to provide faster and more efficient cognitive performance (Kail, 1997; Tamnes et al., 2012; Rolls & Deco, 2015), thus illustrating the potentially complementary roles of grey and white matter to explain differences in cognitive performance.

Beyond the conventional classification of grey and white matter, it is important to consider the heterogeneity within these tissues. Each specific region/tract *within* grey and white matter may offer

unique or shared information in predicting cognitive performance, both within a single metric but also across diverse metrics. Understanding the intricate interplay between these regions/tracts and metrics can provide a more comprehensive understanding of brain-behavior associations.

In summary, there is compelling evidence to suggest that grey and white matter play complementary roles explaining differences in cognitive performance. To date, few studies investigated the combining effects of grey and white matter differences on cognitive performance and so far the literature point towards a complementary roles of the two tissues (Østby et al., 2011; Kievit et al., 2014; Ritchie et al., 2015).

The causes of this paucity are manifold. First, study designs tend to focus on either classical MRI sequences (T1,T2,T2*) or diffusion weighted imaging data. These are sequences with distinct scanner demands and require dedicated analysis expertise, which often leads to papers focusing on one tissue. Moreover, the standard implementation in neuroimaging software, where a brain region/voxel is the *outcome* of a regression equation, makes it considerably more challenging to implement models where multiple brain metrics predict a single phenotypic outcome (e.g. cognitive performance). Finally, the high number of predictors of such models might hinder their feasibility, decrease the statistical power and generate false positives.

To overcome these challenges, we investigated the unique predictive roles of grey and white matter metrics in predicting cognitive performance in childhood. By conducting a multimodal analysis in a regularized structural equation model in a large sample, we overcome many existent weaknesses in previous work, and it allowed us to maximize the likelihood of distinguishing the competing hypotheses of interest. In addition, we will explore the regional distribution of these associations within and across metrics, as well as identify which specific metrics in grey and white matter are more influential in predicting cognitive performance.

This study provides new insights into the complementary information provided by grey and white matter in supporting cognitive performance.

Methods

Participants

The ABCD study (https://abcdstudy.org/) is an ongoing longitudinal study across 21 data acquisition sites enrolling 11 876 children from 9 years old to 16 years old. For more information on ABCD protocols and inclusion/exclusion criteria see Volkow and colleagues (Volkow et al., 2018).

This paper analysed the baseline sample (9-11 years old) from release 4.0 (https://abcdstudy.org/; http://doi.org/10.15154/1523041) that include a sample of 11,876 children. Data entry outliers for 10 participants in the Little Man task were replaced with NA and included in the full information maximum likelihood (FIML) estimation. We included participants with partial data across cognitive tasks and the neuroimaging data in our models.

Given the complexity of the analysis and the a priori challenges associated with high dimensional regularized structural equation modeling approaches, a fully preregistered analysis was not possible. However, to ensure robustness of our findings, we divided the sample into two subsets, allowing us to balance exploratory model optimisation and validation in a non-overlapping sample. In our study, a random sample of 15% of the data were used to optimize the model-building and estimation steps, and the other 85% of the sample was used as a validation sample (Srivastava, 2018) (Figure 1). The set of regions used in the validation sample was derived from the best predictive regions identified in regularized models estimated using the model-building sample (Schwarz, 1978) (the regularisation techniques will be explained more in the Experimental Design and Statistical Analysis section). This strategy allows us to

jointly optimise robustness and flexibility in cases where model estimation, adaptation and convergence are non-trivial, whilst maintaining sufficient power in the validation set to be sufficiently well powered.

The model-building sample consisted of 1,781 children (49.1% female, mean age=9.9, SD=0.6, range=8.9-11.1) and the validation sample included 10,095 children (47.6% female, mean age=9.9, SD=0.6, range=8.9-11).

Cognitive performance

In our modelling approaches, we want to focus on construct capturing a broad sampling of cognitive ability. To this end, we use the same procedure as identified in Sauce et al. (2022), by selecting five cognitive tasks from the NIH Toolbox Cognition Battery were selected: Picture Vocabulary, Flanker, Oral Reading Recognition, Rey Auditory Verbal Learning and Little Man (Sauce et al., 2022). They measure respectively language vocabulary knowledge, attention and inhibitory control, reading decoding skill, verbal learning and memory, visuospatial processing flexibility and attention. All of the tasks were administered using an iPad with support or scoring from a research assistant where needed. For more information on each task, see Luciana et al. (2018).

Next we specified a confirmatory factor model which posits a single latent factor for cognitive performance, reducing measurement error and increasing precision. It enables the model to compute the common variance between these five tasks and thus to have a more accurate representation of cognitive performance (Figure 1B). The measurement model has already been validated in another study, for more information on the rationale for the choice of cognitive tasks and the measurement model see Sauce and colleagues (Sauce et al., 2022). Note, the model chosen here should not be interpreted as a commitment to a strong 'causal g' (see Kievit et al., 2017), but rather as a way to specify a broad cognitive factor which will maximize our statistical power to capture the patterns of interest. A fruitful avenue of future work

will be to examine the degree (or lack of) overlap in the neural mechanisms predicting each individual cognitive domain – a goal beyond the present paper.

Brain structure measures

MRI and DTI scans were collected across the 21 research sites using Siemens Prisma, GE 750 and Philips 3T scanners. Scanning protocols were harmonized across sites and scanners. Full details of all the imaging acquisition protocols and the processing methods used in ABCD are outlined elsewhere (Casey et al., 2018; Hagler et al., 2019).

To determine the importance of grey and white matter, we chose six structural metrics available in the ABCD study: cortical thickness (CT), surface area (SA) and volume (GMV) for grey matter and fractional anisotropy (FA), mean diffusivity (MD) and volume (WMV) for white matter.

Grey Matter

Grey matter measures were estimated from MRI scans with FreeSurfer (version 7.1.1).

The processing steps involved cortical surface reconstruction, subcortical segmentation, removal of non-brain tissue, Talairach transformation, segmentation of white matter and deep gray matter structures, intensity normalization, and surface deformation. The images were registered to a spherical atlas based on individual cortical folding patterns, and the cerebral cortex was parcellated into 34 regions per hemisphere with the Desikan-Killiany atlas (Desikan et al., 2006). The full pipeline can be found elsewhere (Hagler et al., 2019).

The FreeSurfer output includes volume, cortical thickness and surface area (Casey et al., 2018). Considering the challenges of dimensionality in our SEM, we decided to average every region of interest bilaterally (i.e. the brain was parcellated in 34 regions across the brain for each metric).

White Matter

White matter structure can be measured through diffusion-weighted imaging (DWI) and structural MRI (Wandell, 2016). DWI is a technique that allows researchers to record the diffusion of water in the brain using diffusion tensor imaging (DTI) to model the diffusion within the axons and myelin sheath, thus to analyse the directionality of the white matter tracts (Le Bihan, 2003; Assaf & Pasternak, 2008). Structural MRI provides a measure of volume while DTI computes fractional anisotropy, mean diffusivity, radial diffusivity and axial diffusivity (AD).

In the ABCD study, diffusion measures were obtained using tabulated diffusion MRI data from a high angular resolution diffusion imaging (HARDI) sequence with multiple b-values. The images were corrected for head movement, eddy current distortions, B0 distortion, and gradient nonlinearity distortion. The resulting diffusion measures were used for subsequent analyses. The processing steps are described elsewhere (Hagler et al., 2019).

White matter measures include white matter volume, fractional anisotropy and mean diffusivity. The tracts were divided with the AtlasTrack atlas which created 37 regions of interest (17 regions per hemisphere and 3 regions aside) (Hagler et al., 2019). We followed a similar procedure as with the grey matter metrics and averaged every tract of interest bilaterally.

Experimental Design and Statistical Analysis

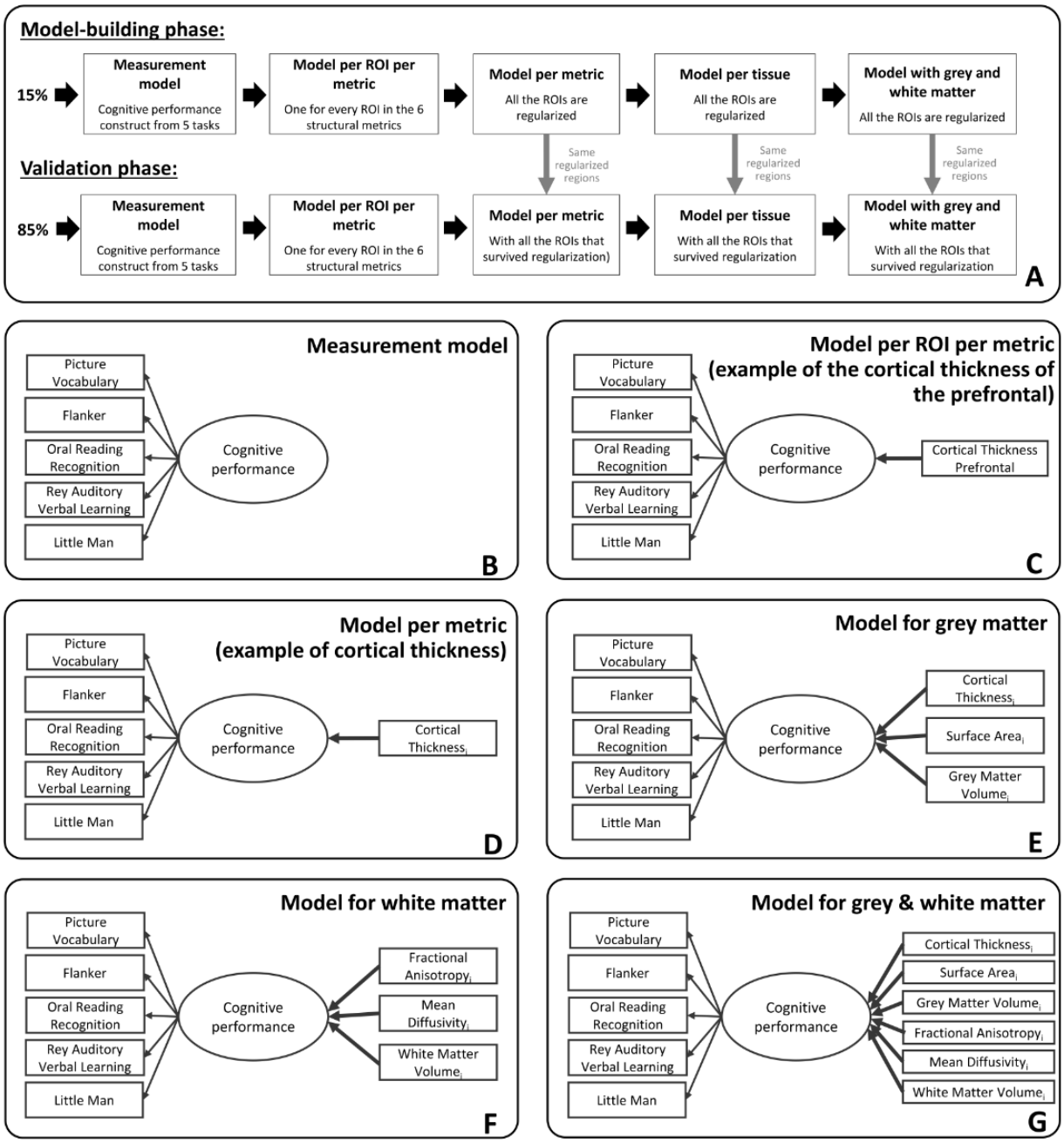


Figure 1: (A) Steps followed during the model-building and the validation phases of the study. The regions of interest fitted into the models of the validation phases resulted from an initial regularisation made in the model-building phase. (B) Measurement model of cognitive performance based on five cognitive tasks. (C) Example of a MIMIC model path diagram of a model per ROI per metric where the one metric of one region of interest (cortical thickness of the prefrontal) predict cognitive performance. (D) Example of a MIMIC model path diagram of a model per metric where

all regions of interest of one metric (here cortical thickness) that survived regularization predict cognitive performance. (E) MIMIC model path diagram for the grey matter model with cortical thickness, surface area and grey matter volume predicting cognitive performance. i represents the regions of interest from each metric in grey matter that survived regularization in the model and can go from 1 to 34 (as we use the Desikan-Killiany atlas and average bilaterally). (F) MIMIC model path diagram for the white matter model with fractional anisotropy, mean diffusivity and white matter volume predicting cognitive performance. j represents the regions of interest from each metric in white matter that survived regularization in the model and can go from 1 to 20 (as we use the AtlasTrack atlas and average bilaterally). (G) Path diagram for the grey and white matter model with cortical thickness, surface area, grey matter, fractional anisotropy, mean diffusivity and white matter volume of the regions that survived regularization predicting cognitive performance.

Next, we specified the core questions investigated by our analyses.

- *Do the metrics of grey matter and white matter demonstrate complementary roles predicting cognitive performance?*
- *Do the different metrics in each tissue have unique predictive roles? If so, which metric is most important to predict cognitive performance?*
- *For each metric, do the different regions of interest have unique predictive roles?*
- *Do the same regions of interest have the strongest predictive power across different metrics in each tissue?*

Our focus will be on both the tissue scale (grey and white matter) and the metrics scale (cortical thickness, surface area, grey matter volume, fractional anisotropy, mean diffusivity and white matter volume). For each scale, we will study if the predictors of cognitive performance each have a unique role (i.e. there is no overlap in the information contributed by the different variables to predict cognitive performance) or if they have complementary roles (i.e. there is a partial or total overlap in the information contributed by the different variables to predict cognitive performance).

158 The study uses structural equation model approach to evaluate the hypothesis that brain structural
159 metrics predict the latent variable, cognitive performance. Structural equation models, and more
160 particularly MIMIC models (Multiple Indicator Multiple Cause) offer an effective way to model cognition
161 as a latent variable and to estimate the contribution of multiple simultaneous hypothesized causes to
162 explain individual differences in cognitive performance (Jöreskog & Goldberger, 1975; Kievit et al., 2014).

163 First, we estimated a series of structural equation models to test how each metric in each individual
164 region/tract predicted cognitive performance (i.e., every region in all the six metrics have been computed
165 independently). The key question of interest is the strength of the key parameter highlighted in bold
166 (Figure 1C).

167 Next, to assess how each metric predicted cognitive performance we used a regularized structural
168 equation modelling (SEM) approach (Jacobucci et al., 2019) which incorporates a penalty on key
169 parameters of interest. Specifically, it allows us to have many regions/metrics simultaneously predict the
170 outcome, with a penalty on the path estimates from brain metrics to the cognitive latent variable, that
171 induces sparsity. Regularization is a method that imposes a penalty in order to decrease the complexity
172 of the model while keeping the variables that are the most important in predicting the outcome. For
173 instance, it allows us to include the cortical thickness measures from all 34 regions of interest as
174 simultaneously predicting cognitive performance with a lasso penalty which pushes parameter estimates
175 of small or absent effects to 0, and retains only those regions which contribute meaningfully in predicting
176 the outcome for the validation sample. This accounts for the partial overlap in variance arising from the
177 inherent way the metrics are measuring the same underlying structure. For instance, grey matter volume
178 is a product of surface and cortical thickness. If grey matter volume functions as a deterministic
179 transformation with no additional information beyond cortical thickness or surface area for cognitive
180 performance, our regularization approach should ensure all redundant parameter estimates are pushed
181 to 0. For example, if the cortical thickness, surface area and grey matter volume of the frontal pole each

182 have large effects on cognitive performance and these effects are redundant (they include the same
183 information to predict cognitive performance), then the measures with the smaller effects of the three
184 will be pushed to 0.

185 We developed a procedure to select a subset of regions/tracts based on their predictive ability in the
186 model-building sample (e.g., 15% of the total sample). The output of each regularized model estimation
187 was a set of all the regions/tracts with a regularized beta different from zero, or considered 'important'
188 in a regularized framework. These regions were entered into the models predicting cognitive performance
189 in the validation sample (e.g., 85% of the total sample). We implemented this process across six models
190 per metric (Figure 1D), two models per tissue (Figures 1E-F) and the model combining grey and white
191 matter metrics (Figure 1G). The benefit of this approach is to have a parsimonious representation of the
192 key regions which help predict cognitive performance in each metric. We also examined 'full' models
193 combining every region of interest within a metric (e.g. the 34 regions for cortical thickness), within a
194 tissue (e.g. the 102 regions for the three grey matter metrics and the 60 regions for the three white matter
195 metrics) and within grey and white matter metric (e.g. the 162 regions across the six metrics).

196 To compare the predictive information of grey and white matter, three models were fitted: one model
197 with the regions extracted from the regularization of the three grey matter metrics, one model with the
198 regions extracted from the regularization of the three white matter metrics and a final model with the
199 regions extracted from the regularization of the grey and white matter metrics (e.g., as displayed in Figure
200 1A, the regions were selected because they survived the regularisation in the model-building sample). For
201 each model, we use a likelihood ratio test to examine whether the inclusion of a tissue (grey/white matter)
202 or a metric within a tissue (e.g. cortical thickness) improves the model compared to a model where the
203 paths corresponding to an additional metric/tissue are constrained to 0. For each final model, we
204 extracted the (adjusted) R-square to assess the proportion of variance in cognitive performance explained
205 by each model.

From the analyses, we can imagine our results being captured by one of three (simplified) scenarios:

- Grey and white matter metrics give the same, non-complementary information to predict cognitive performance. The models with only grey matter, only white matter or both will have a similar adjusted R-squared (Figure 2A), and model selection would favour a model with only one tissue.
- Grey and white matter metrics give fully distinct/complementary information to predict cognitive performance. Under this scenario, model selection would favour a model including both tissues and the joint R-squared would approximate the sum of the R-squared of each tissue in isolation (Figure 2B).
- Grey and white matter metrics give complementary, but partially overlapping, information to predict cognitive performance. In this case, model selection would favour a model with both tissues, and the joint R-squared will be higher than the one of the most predictive tissue, but lower than the sum of the R-squared of each tissue in isolation (i.e. the R-squareds are neither interchangeable nor fully additive) (Figure 2C).

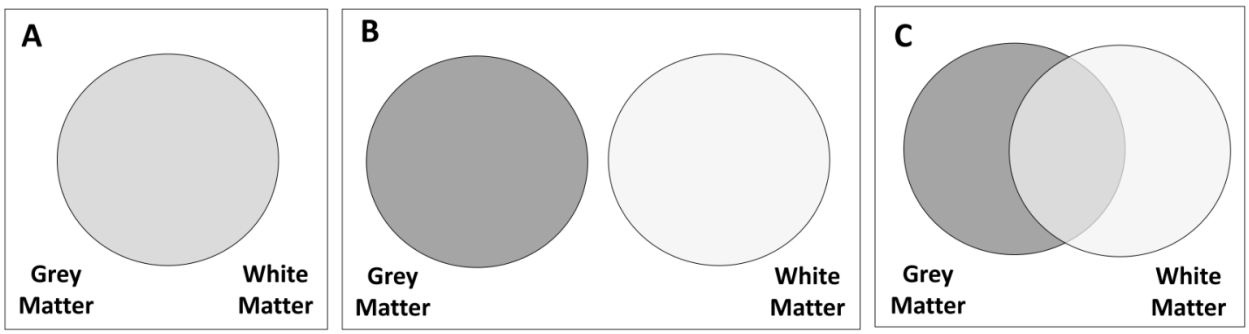


Figure 2: Hypotheses regarding the information provided by grey and white matter metrics. (A) Redundant, grey and white matter give the same information. (B) Independent, grey and white matter give completely different information. (C) Partially overlapping, grey and white matter give both similar and different information.

These analyses were replicated separately for male and female participants to examine the potential impacts of sex differences in brain structure for the findings. In an additional analysis, total intracranial volume (TIV) was added as a third predictor to explain the differences in cognitive performance. Accounting for TIV in the model is a contentious issue (Hyatt et al., 2020), as it modifies the question from whether it is the absolute size (or thickness, anisotropy,..) of a region that matters to predict cognitive performance, or the relative measure given a participant's TIV.

Regularization has many strengths, as outlined before, but this comes at a specific cost, namely its attempt to build a simplified model. In other words, if two regions have almost the same predictive power, but are somewhat redundant in this regard, a regularized model is likely to push the parameter estimate of one of the two regions to 0 and 'keep' the other region instead. If we were to report the regional pattern as such, this may lead to an overinterpretation of the absence and presence of some regions. To examine the regional specificity of our findings, we conducted a repeated regularization procedure, iterated 100 times using a different 15% sample on each occasion. Across these 100 iterations, we computed the percentage of times each region survived regularization (i.e. when a given region had a non-zero estimate in the final model). This approach allowed us to assess the stability of each region/tract in providing unique information to explain variations in cognitive performance. Within each model (see Figure 1D-G), we calculated the percentage of instances in which a region/tract survived regularization. Regions/tracts with higher percentage values demonstrate a more consistent and distinct contribution to the prediction of cognitive performance. In the model employing a single metric (Figure 1D), this reflects the differential information provided by distinct regions/tracts within the same metric. In the models incorporating multiple metrics (Figure 1E-G), the likelihood of regions/tracts surviving regularization is lower if they provide redundant information across different metrics. Therefore, regions/tracts with a higher

247 percentage not only demonstrate a unique predictive role across different regions/tracts but also across
248 the various metrics.

249 We used the following guidelines to assess the good fit of the models: root mean square error of
250 approximation (RMSEA) <0.05 (acceptable: 0.05–0.08), comparative fit index (CFI)>0.97 (acceptable:
251 0.95–0.97) and standardized root mean square residual (SRMR)<0.05 (acceptable: 0.05–0.10)
252 (Schermelleh-Engel et al., 2003; Mueller & Hancock, 2008). We compared models fit using akaike
253 information criterion (AIC) and bayesian information criterion (BIC). These parameters penalize models
254 with a higher number of predictors, thus encouraging the selection of more parsimonious models that
255 explain the data well while avoiding overfitting.

256 All analyses were carried out on the data of the first wave using R, version 4.1.0 (http:// www.r-
257 project.org/) and the lavaan package (Rosseel, 2012). All models were fit using Maximum Likelihood
258 Estimation, with FIML to account for missing data and robust estimation with adjusted standard errors to
259 deal with deviations from (multivariate) normality.

260 Data and Code Accessibility

261 Data can be requested through [https://nda.nih.gov/], and the code to reproduce our analyses is available
262 on [https://osf.io/ryskf/].

263

264 **Results**

265 Measurement model (validation, parameters, invariance)

266 To assess cognitive performance, we used the same measurement model built by Sauce and colleagues
267 (Sauce et al., 2022) in a slightly different sample. Model estimates were highly similar.

The confirmatory factor model fitted the data well in the validation sample, x^2 = 156.404, degrees of freedom (df)= 5, p<0.001, RMSEA = 0.055 (0.00 - 0.070), CFI= 0.980, SRMR= 0.020. This result demonstrates that the common variance among all five cognitive tasks can be captured by one latent variable that we call here "cognitive performance".

Oral Reading Recognition task and Picture Vocabulary task have the strongest standardized factor loadings (0.74 and 0.7 respectively). Rey Auditory Verbal Learning task, Little Man task and Flanker task are mildly predicted by the construct between 0.41 and 0.53. (Figure 3). The results show little change when we add age as a predictor of the latent variable.

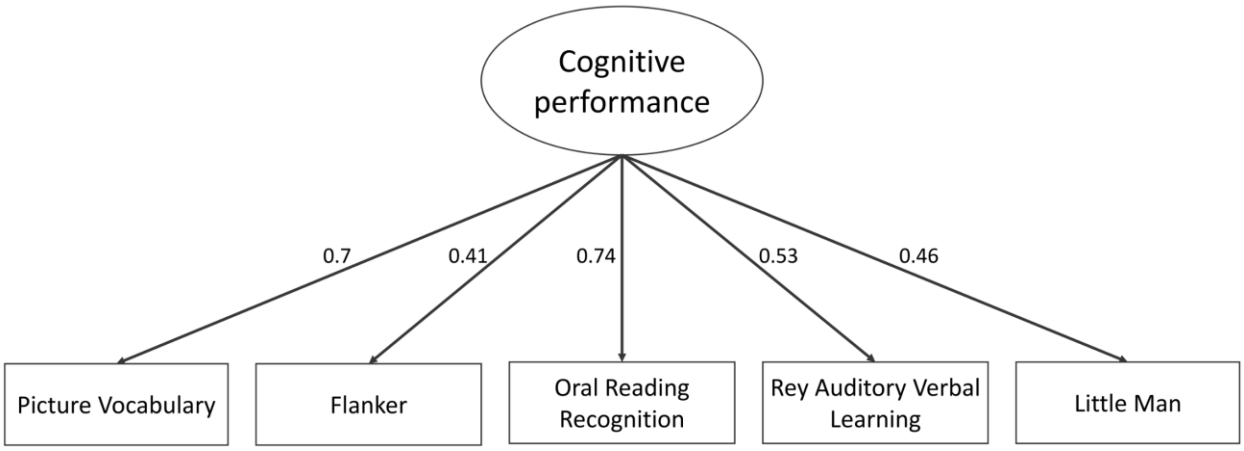


Figure 3: Path diagram of a measurement model of cognitive performance based on five cognitive tasks. All variables are standardized and significant.

1. <u>Grey and white matter give both overlapping and unique information to predict cognitive performance</u>

We fitted the full model including all regions/tracts that survived regularisation in grey & white matter metrics, and compared it to a model which includes the same predictors, but constrains either all grey or

284 all white matter predictors to 0. This comparison allows us to test if the metrics in grey matter and white
285 matter have complementary roles predicting cognitive performance.

286 The model with both grey and white matter fitted the data well (x^2 = 793.903, df= 205, p<0.001, RMSEA =
287 0.017 (0.016 - 0.019), CFI= 0.935, TLI=0.918 SRMR= 0.011) and showed the best performance among the
288 three models (AICdiff > 265 , BICdiff > 136 in favour of the model with both grey and white matter), and
289 explained 19.0% of the variance in cognitive performance. The model with only grey matter explained
290 15.4% of the variance and the one with only white matter explained 12.4% of the variance (Table 1).

291 These findings demonstrate that grey and white matter bring both overlapping and unique information
292 to predict the differences in cognitive performance in line with hypothesis C – 'partially overlapping' – in
293 Figure 2.

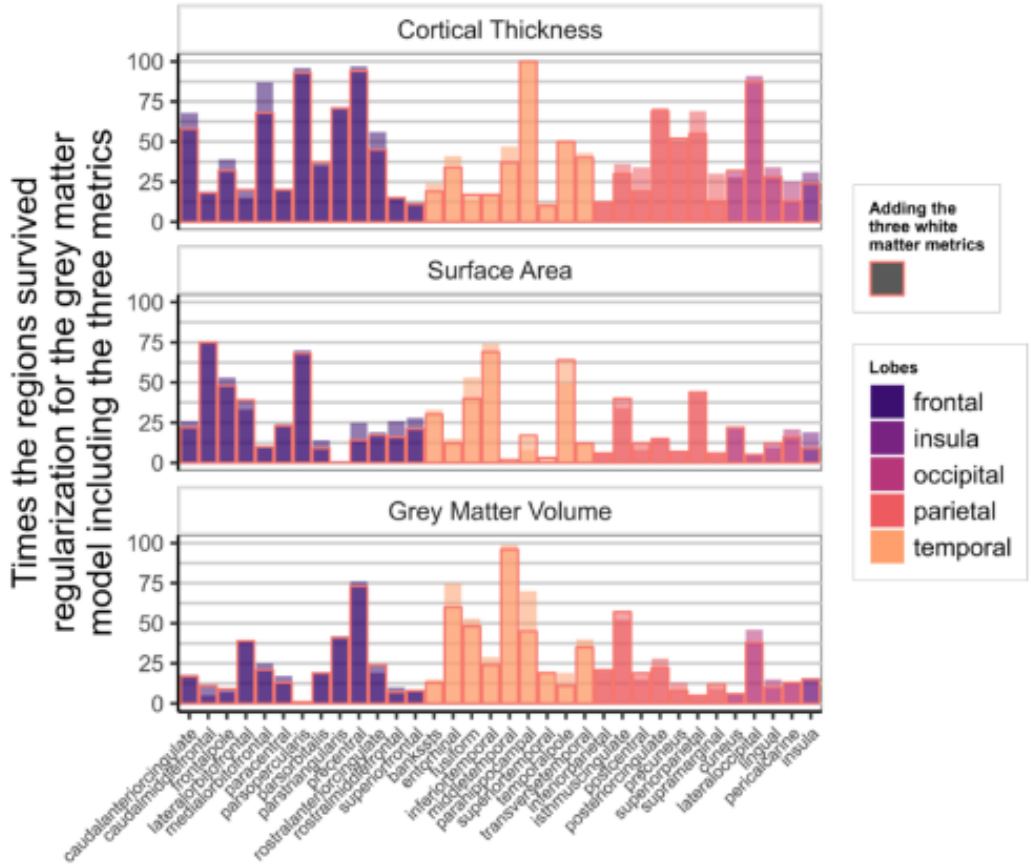
294 These results were replicated in samples comprising exclusively of male (R-square adjusted for grey
295 matter and white matter= 17.4%, for grey matter= 14.6% and for white matter= 10.8%) and also female
296 participants (R-square adjusted for grey matter and white matter= 23.0%, for grey matter= 17.8% and for
297 white matter= 15.9%).

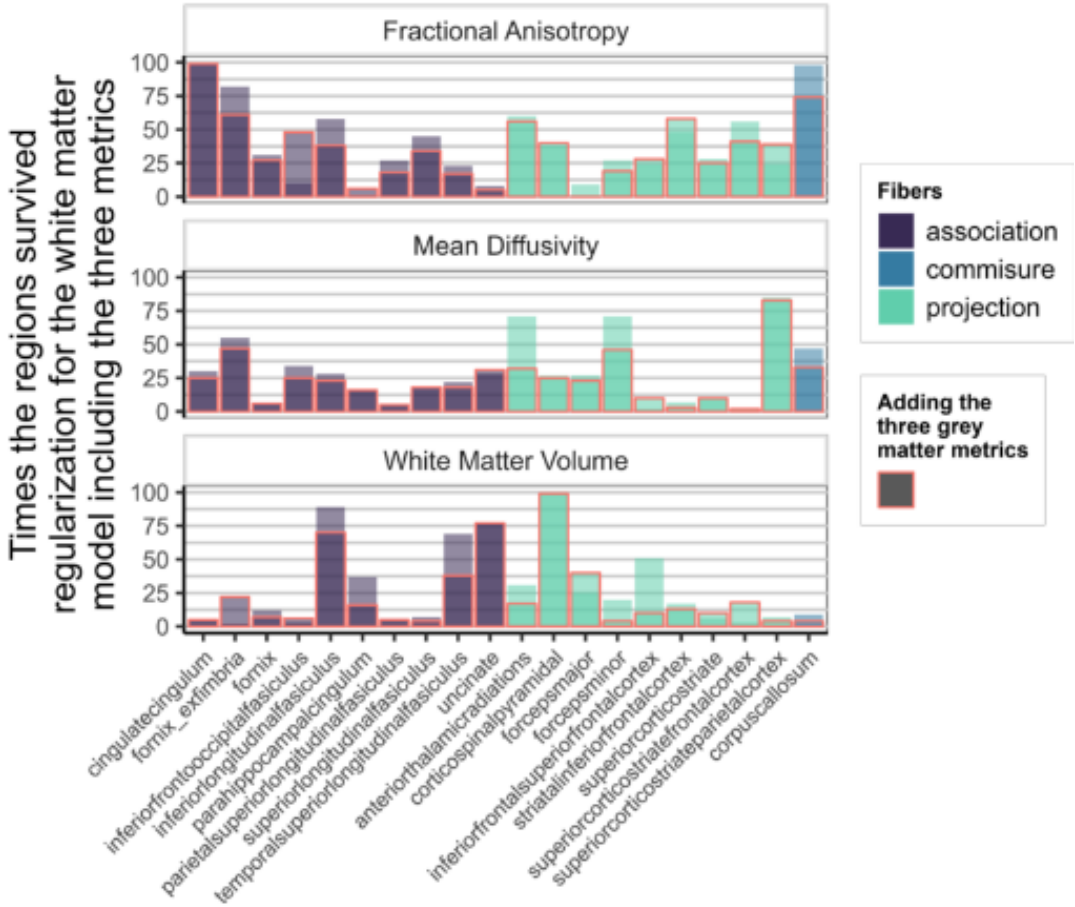
298 In the model accounting for total intracranial volume, we also compared a model with only TIV as a
299 predictor and a model with grey matter, white matter and TIV as a predictor instead of only grey and
300 white matter. The model with grey matter, white matter and TIV explained the best the variance in
301 cognitive performance (R-square adjusted for grey matter and white matter and TIV= 19.1%), while the
302 model with only TIV explained 6.8%. The addition of TIV as a predictor did not improve the fit of the model,
303 nor did its inclusion substantially altered the parameters estimates of the regions/tracts.

304

	Grey Matter model	White Matter model	Grey and White Matter model
R-square (adjusted)	0.154	0.124	**0.190**
AIC	122941.1	123153.5	**122675.3**
BIC	123277.4	123389.6	**123140.5**

Table 1: Estimates of the models fit and the variance explained by the predictors within each model. The best model is shown in bold.


Times the regions survived regularization for the grey matter model including the three metrics
Cortical Thickness
Surface Area
Grey Matter Volume
0
25
50
75
100
caudalanteriorcingulate
caudalmiddlefrontal
frontalpole
lateralorbitofrontal
medialorbitofrontal
paracentral
parsopercularis
parsorbitalis
parstriangularis
precentral
rostralanteriorcingulate
rostralmiddlefrontal
superiorfrontal
bankssts
entorhinal
fusiform
inferiortemporal
middletemporal
parahippocampal
superiortemporal
temporalpole
transversetemporal
inferiorparietal
isthmuscingulate
posteriorcingulate
precentral
precuneus
supramarginal
cuneus
lateraloccipital
lingual
pericalcarine
insula
Adding the three white matter metrics
Lobes
frontal
insula
occipital
parietal
temporal


Times the regions survived regularization for the white matter model including the three metrics
Fractional Anisotropy
Mean Diffusivity
White Matter Volume
0
25
50
75
100
cingulatecingulum
fornix_exfimbria
fornix
inferiorfrontooccipitalfasiculus
inferiorlongitudinalfasiculus
parahippocampalcingulum
parietalsuperiorlongitudinalfasiculus
superiorlongitudinalfasiculus
temporalsuperiorlongitudinalfasiculus
uncinate
anteriorthalamicradiations
corticospinalpyramidal
forcepsmajor
forcepsminor
inferiorfrontalsuperiorfrontalcortex
striatalinferiorfrontalcortex
superiorcortex
superiorcorticostriatefrontalcortex
superiorcorticostriateparietalcortex
corpuscallosum
Fibers
association
commisure
projection
Adding the three grey matter metrics

307

Figure 4: Percentage of times one region/tract in a metric survived regularization in the model with the three metrics of one tissue (top: grey matter, bottom: white matter) (Model E & F in Fig 1). The columns highlighted in red represent the percentage of survival for the region when the model includes the metrics of the other tissue (Model G in Fig 1)

Figure 4 illustrates the overlap in information within grey and white matter metrics, as evident from the depicted contrast in the percentage of times one region/tract is consistently surviving regularization in a model with only the metrics on one tissue versus a model including the metrics of both grey and white matter. Despite this overlap, certain regions still reliably predict cognitive performance, highlighting the distinct information observed within both tissue types.

2. <u>Some metrics are more predictive of cognitive performance than others</u>

Next, we investigated if the different metrics in each tissue have a unique predictive role, and thus if the choice of the metric was important when you want to predict cognitive performance. Comparing the predictive power of the different metrics, we plotted the standardized estimate model parameters of every region across metrics. Figure 5 shows the range of observed values in the path estimates for each metrics, within the grey matter metrics, surface area (range β_{std} = [0.092;0.265]) and volume (range β_{std} = [0.123;0.275]) were overall stronger predictors of cognitive performance compared with cortical thickness (range β_{std} = [-0.06;0.145]). This is somewhat surprising given the prominence of cortical thickness as the metric of choice in (developmental) cognitive neuroimaging studies of individual differences. For instance, since 2016, the number of papers using cortical thickness is 5 times greater than the number of papers using grey matter volume (N=4808 for cortical thickness and N=823 for grey matter volume in a PubMed search as of 16/09/2022).

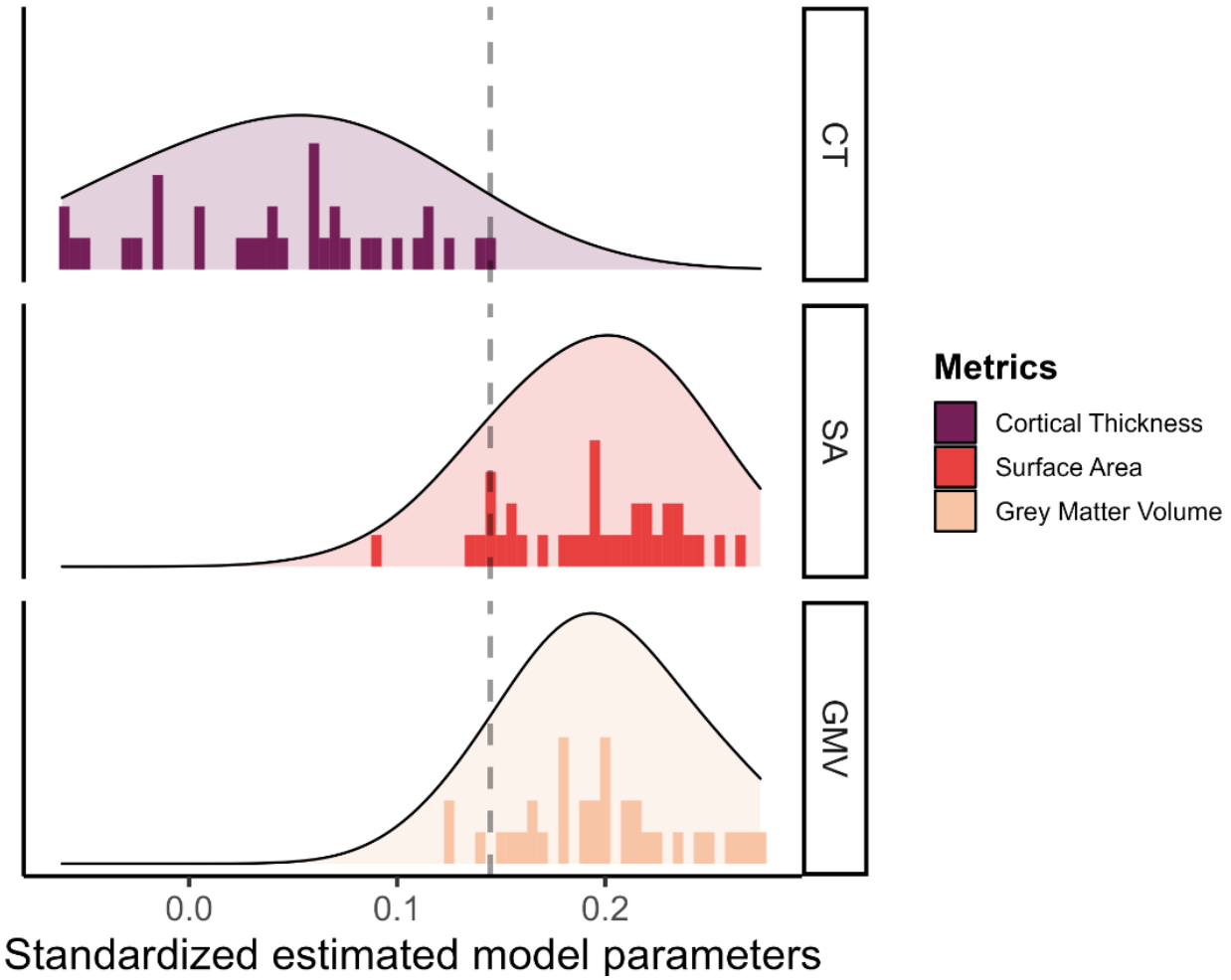


Figure 5: Standardized estimated model parameters on how one region of interest in one metric predicts cognitive performance in the models per ROI per metric for the three grey matter metrics.

The results for the white matter metrics is represented in Figure 6. Surprisingly, white matter volume is overall the strongest predictor of cognitive performance (rangeβ_{std} = [0.169;0.303]), followed by fractional anisotropy (range β_{std} = [-0.010;0.142]) and mean diffusivity (range β_{std} = [-0.062;0.031]).

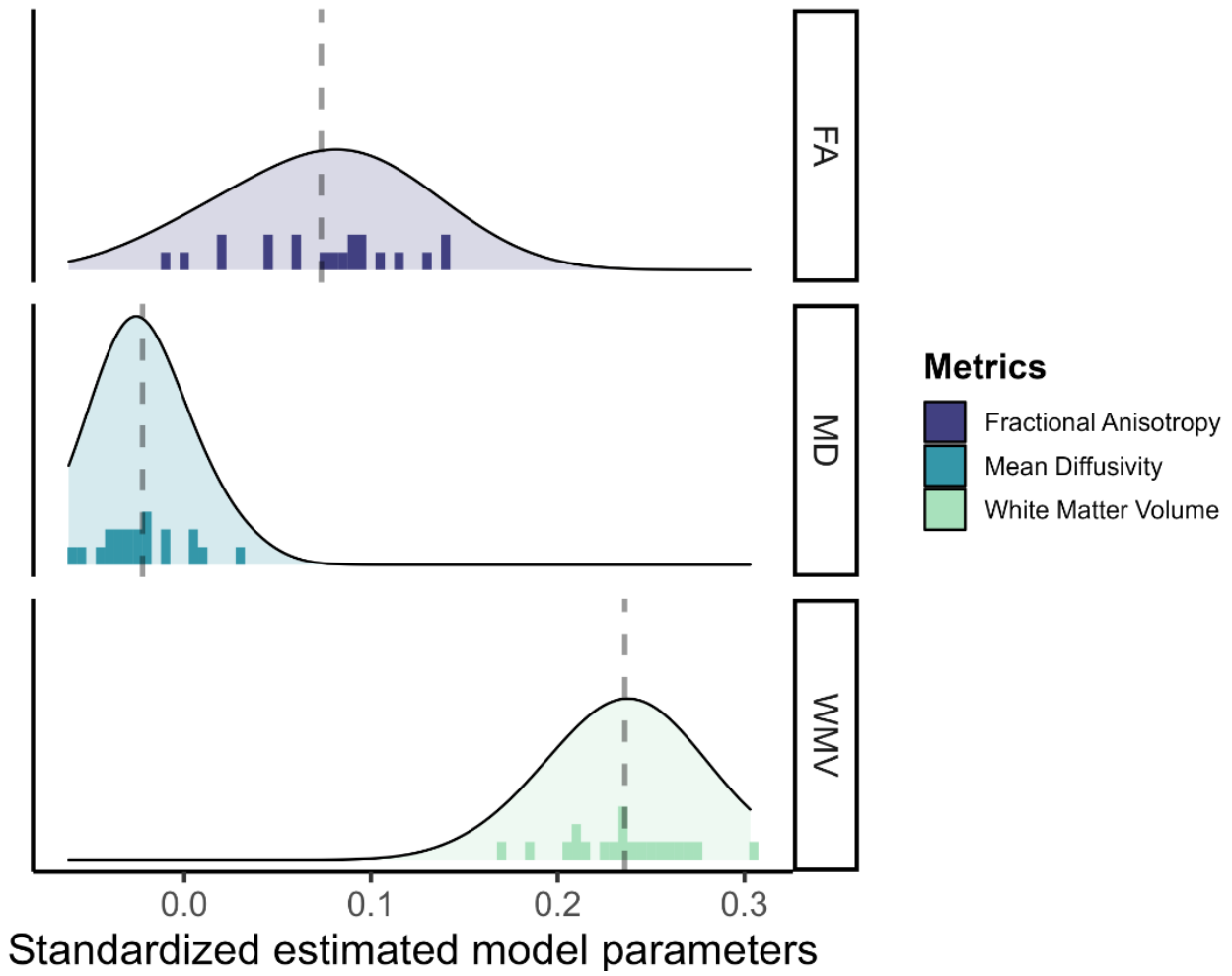


Figure 6: Standardized estimated model parameters on how one region of interest in one metric predicts cognitive performance in the models per ROI per metric for the three white matter metrics.

These findings were replicated in both male and female participants.

These results provide evidence that different metrics within grey and white matter better predict cognitive performance.

3. <u>Within a metric, not every region gives the same information to the prediction of cognitive performance</u>

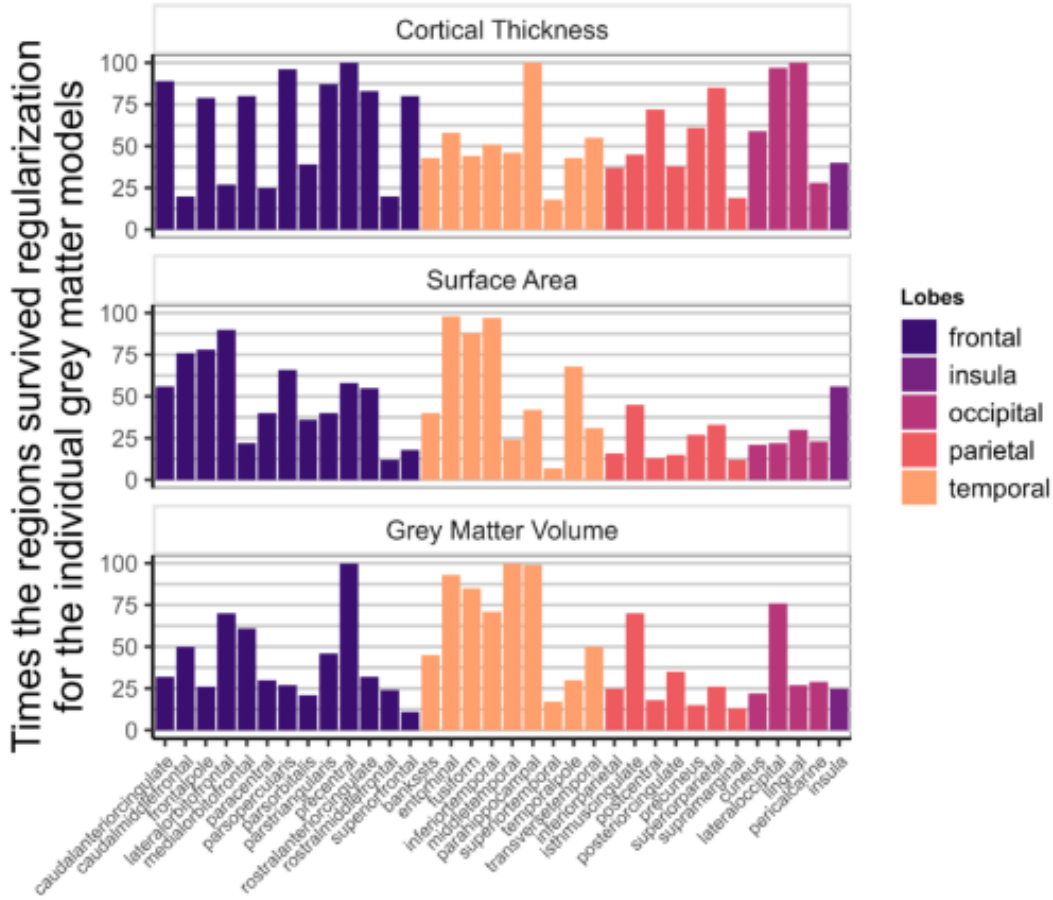
In addition to different metrics showing distinct predictive strengths as observed above, it is also highly plausible (e.g. Basten et al., 2015) that there is regional specificity.In other words, that different regions of interest might have a unique predictive role, and that this role depends on the metrics being studied. To assess the effects of the different regions within a metric, we compared a model with freely estimated parameters to a model where all the parameters are equality constrained, which captures the hypothesis that each region contributes equally. With three exceptions, the free models showed the best fit to the data (AICdiff > 50 , BICdiff > 15 in favour of the models with freely estimated parameters for 15/18 of the models with all the regions and the ones with the regularized regions). These exceptions pertained exclusively to BIC criteria in the models with all the regions, their occurrence can be attributed to the penalization of increased model complexity.

In addition, the regularized models, which favour sparse models with only few predictors, always retained multiple regions even within the same metric. If a single 'key' region contains all relevant predictive information, we would not expect to observe this pattern. Across the nine models that underwent regularization, each model estimation showed at least 30% of all regions to have a regularized beta different from zero, demonstrating the regional specificity and complementarity hypothesized above.

Figure 7 shows how often each region appeared in the final, regularized models across 100 different subsets of the data. Regions or tracts with a high percentage of survival are those that most consistently provide unique information to predict differences in cognitive performance compared to all other different regions and tracts within the metric. For instance, a researcher interested in the parietal cortex will see in Figure 7 that regions within this lobe are more consistently important (compared to regions in other lobes) in a model including only cortical thickness, than in a model that analyses surface area or grey matter volume.

366 These converging findings demonstrate that to have a better picture of cognitive performance we would

367 need to assess different regions in a metric simultaneously.


Times the regions survived regularization for the individual grey matter models
Cortical Thickness
Surface Area
Grey Matter Volume
100
75
50
25
0
caudalanteriorcingulate
caudalmiddlefrontal
frontalpole
lateralorbitofrontal
medialorbitofrontal
paracentral
parsopercularis
parsorbitalis
parstriangularis
precentral
rostralanteriorcingulate
rostralmiddlefrontal
superiorfrontal
bankssts
entorhinal
fusiform
inferiortemporal
middletemporal
parahippocampal
superiortemporal
temporalpole
transversetemporal
inferiorparietal
isthmuscingulate
postcentral
posteriorcingulate
precuneus
superiorparietal
supramarginal
cuneus
lateraloccipital
lingual
pericalcarine
insula
Lobes
frontal
insula
occipital
parietal
temporal

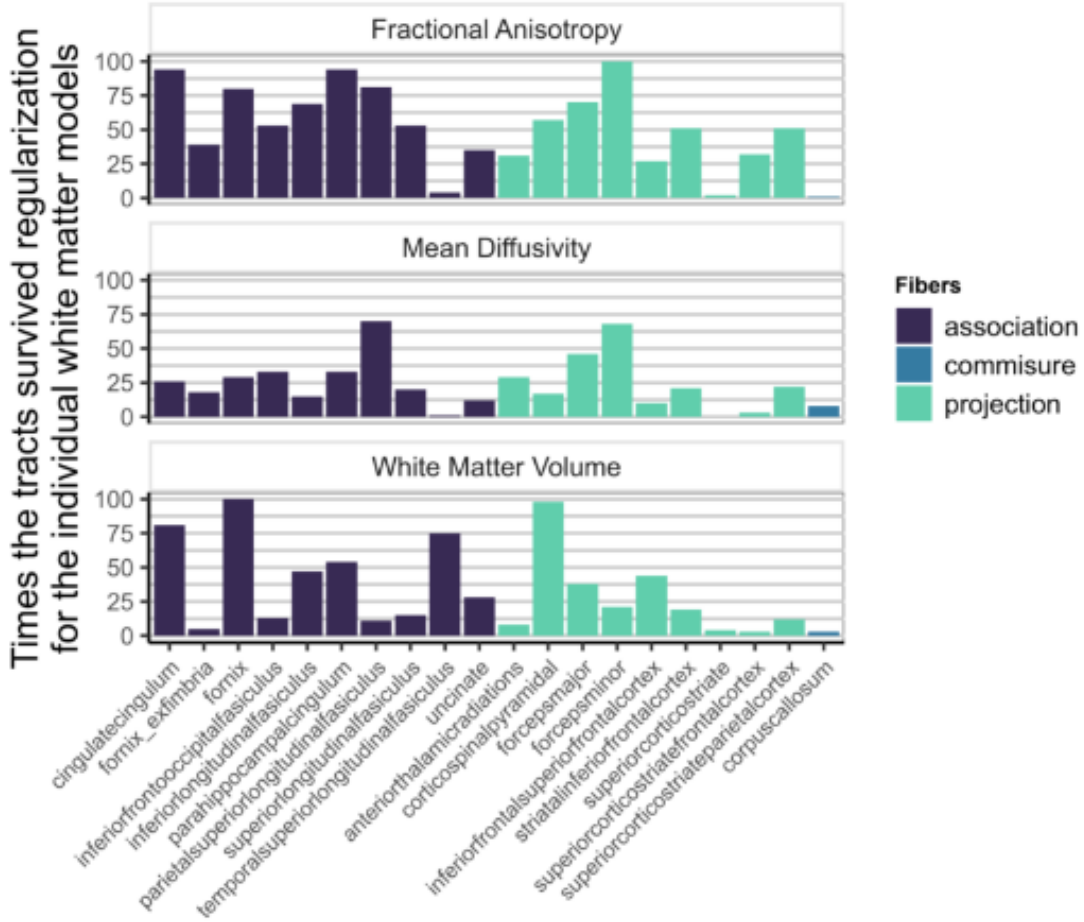

Times the tracts survived regularization for the individual white matter models
Fractional Anisotropy
Mean Diffusivity
White Matter Volume
100
75
50
25
0
cingulatecingulum
fornix_exfimbria
fornix
inferiorfrontooccipitalfasiculus
inferiorlongitudinalfasiculus
parahippocampalcingulum
parietalsuperiorlongitudinalfasiculus
superiorlongitudinalfasiculus
temporalsuperiorlongitudinalfasiculus
uncinate
anteriorthalamicradiations
corticospinalpyramidal
forcepsmajor
forcepsminor
inferiorfrontalsuperiorfrontalcortex
striatalinferiorfrontalcortex
superiorcorticostriate
superiorcorticostriatefrontalcortex
superiorcorticostriateparietalcortex
corpuscallosum
Fibers
association
commisure
projection

Figure 7: Percentage of times individual region/tract in a metric survived regularization in the individual models for each metric (Model D in Fig 1)

4. <u>Across metrics, different regions give unique information to the prediction of cognitive performance</u>

Finally, we investigated if the same regions of interest bring unique information to predict cognitive performance across different metrics in each tissue (e.g. if the fractional anisotropy and white matter in a specific tract contain similar information then the regularization will likely regularize one of the metric to 0 for that given tract). A lasso regression was estimated with every region of interest in each grey (or white) matter metrics, the remaining regions were entered in a model as predictors of cognitive performance. This regression allows us to check which regions were nonessential to the prediction of cognitive performance when taking into account all the regions (or tracts) in the three metrics.

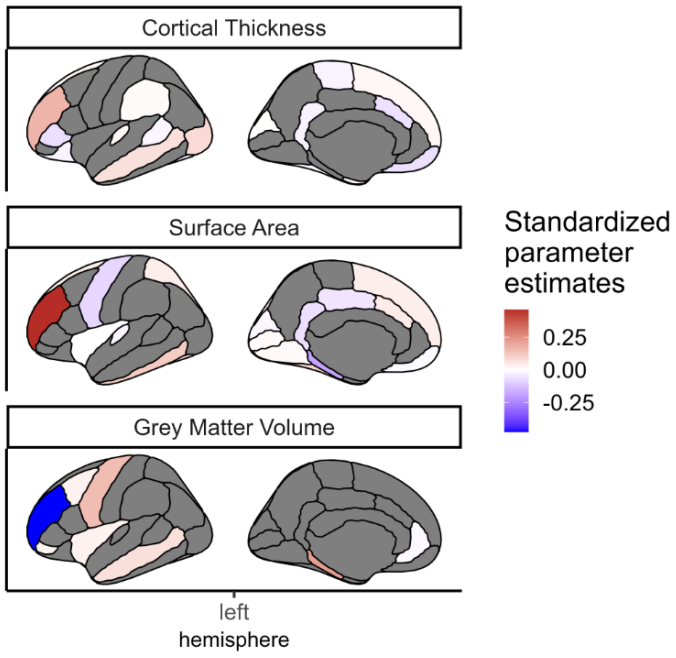
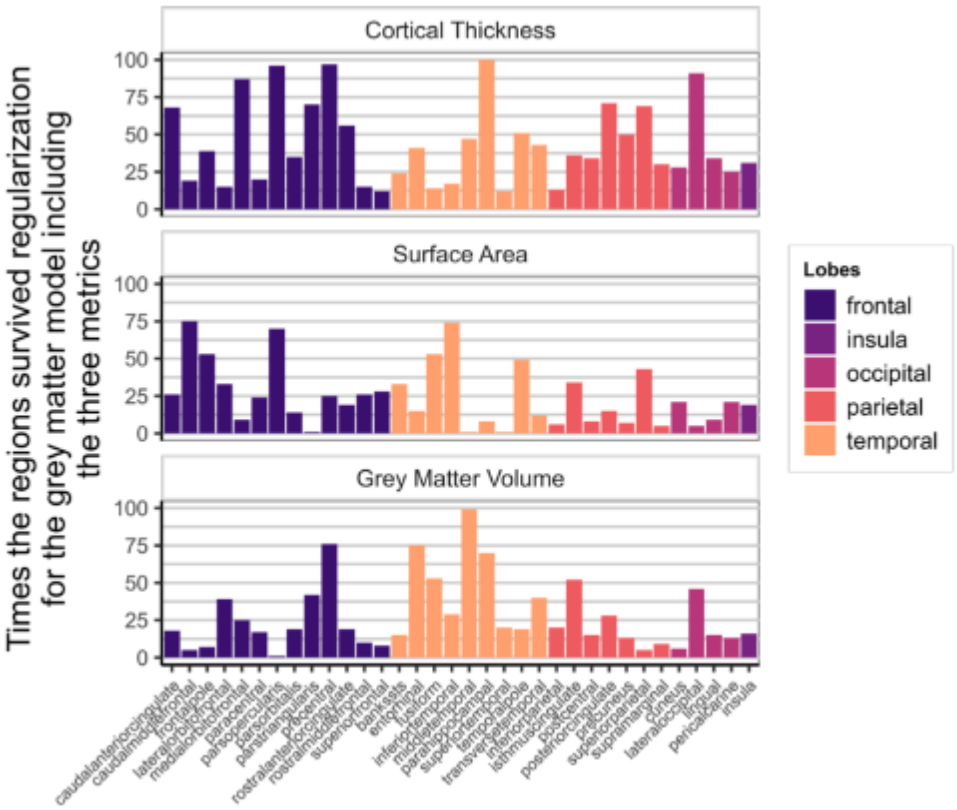
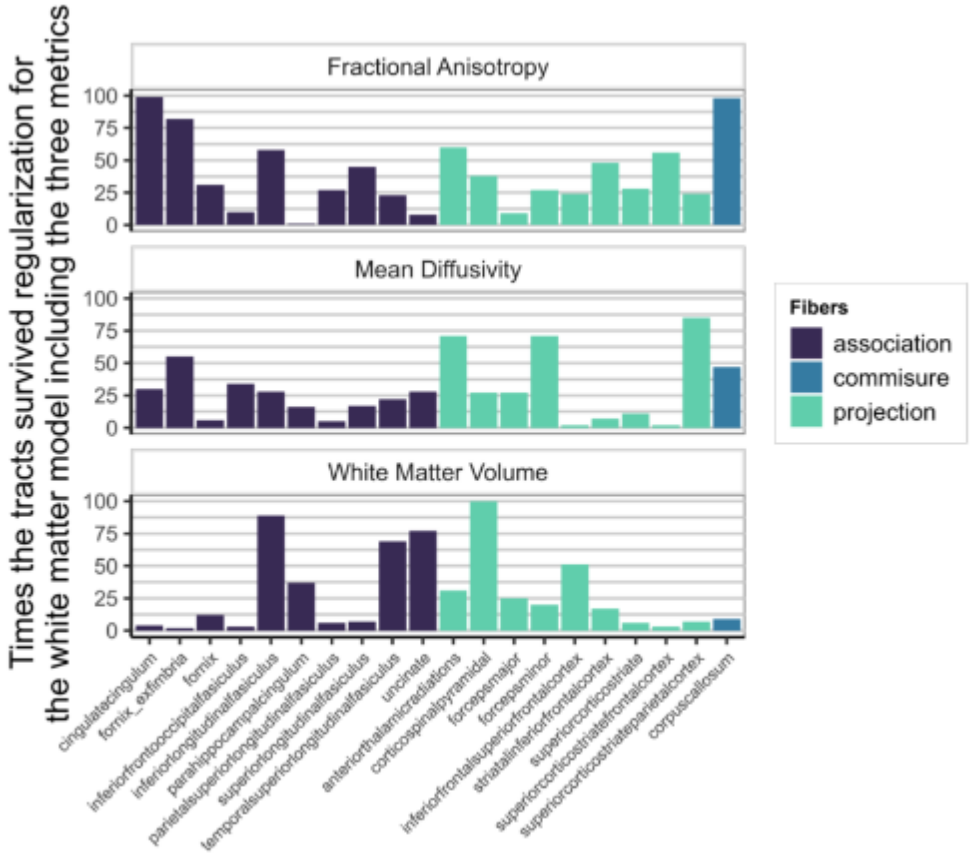


Figure 8: Regions of interest in each metric that survived the regularisation including the three metrics of grey matter. The colour indicated the standardized estimated model parameters

We found that despite the strong penalty included in the regularisations, the model still retained regions from each of the grey matter metrics, as well as each of the three white matter metrics, to be significantly predictive of cognitive performance (Figure 8). This suggests that different regions/tracts give unique information to the prediction of cognitive performance.


Times the regions survived regularization for the grey matter model including the three metrics
Cortical Thickness
Surface Area
Grey Matter Volume
Lobes
frontal
insula
occipital
parietal
temporal


Times the tracts survived regularization for the white matter model including the three metrics
Fractional Anisotropy
Mean Diffusivity
White Matter Volume
Fibers
association
commisure
projection

389

Figure 9: Percentage of times one region in a metric survived regularization in the model with every grey matter metrics included (top figure – Model E in Fig 1) and in the model with every white matter metrics (bottom figure – Model F in Fig 1).

Figure 9 illustrates the survival rates of each region/tract across 100 regularizations in a model incorporating either the three grey matter metrics or the three white matter metrics as predictors of cognitive performance. Comparing the percentage values to those in Figure 7, it becomes evident that a portion of the unique information associated with each region in each metric becomes redundant when performing the regularization analysis with combined grey and white matter metrics.

The complete results of all the models studies in this paper are available on OSF [https://osf.io/eygwz].

Discussion

In a large developmental sample, we examined the role of grey and white matter in supporting cognitive performance. Using regularized structural equation model, we observed that the variance in cognitive performance explained by both grey and white matter (19.0%) is considerably greater than either one in isolation (15.4/12.4%), but not fully additive. This demonstrates that grey and white matter structure bring both unique as well as shared information in predicting individual differences in cognitive abilities in children.

The observed shared information is unsurprising, given the interdependence of these tissues both during development and aging (Giorgio et al., 2010; Hoagey et al., 2019). Previous work suggests close associations between grey and white matter during development (Tamnes, Østby, Fjell, et al., 2010; Kochunov et al., 2011) and some similarity in the ability of grey and white matter measures to predict

412 cognition (Hoagey et al., 2019). Zhao et al. (2021) proposed that differences in sample ages may explain
413 differences in regional and metric specific findings (Zhao et al., 2021) in line with Østby et al. (2011) and
414 Fuhrmann et al. (2020) who observed age-dependent differences in the nature and strength of the
415 relationship between grey matter, white matter and cognitive performance (Østby et al., 2011; Fuhrmann
416 et al., 2020).

417 Regarding the secondary goals, we found evidence that different metrics within grey and white matter
418 have distinct predictive power in this sample. For grey matter, surface area and volume show the best
419 predictive power compared to cortical thickness, a surprising finding given the prevalence of studies which
420 (only) use cortical thickness as a structural predictor of cognitive performance. However, this result is
421 consistent with several studies reporting that surface area is more related to cognitive performance than
422 cortical thickness (Borgeest et al., 2021; Fürtjes et al., 2023; Pulli et al., 2023). Moreover, recent evidence
423 within the same sample suggests that cortical thickness has lower reliability than surface area or volume,
424 which may differentially attenuate the effects a given study observes (Parsons et al., 2023).

425 For white matter, we found that in this sample white matter volume is the strongest predictor compared
426 to more 'advanced' diffusion-based metrics such as fractional anisotropy and mean diffusivity. This is also
427 unexpected since tract integrity measures are seen as more closely associated with the underlying
428 physiology of tracts and their lesions (Fjell et al., 2008; Xing et al., 2021).

429 The regional differentiation found at the level of the metrics and the tissues in our models illustrates the
430 importance of selecting regions of interest. Some regions emerge as robust predictors of variance in
431 cognitive performance, while others contribute minimally or lack unique informational value. However, it
432 is apparent that the predictive efficacy of individual regions/tracts varies across the metrics within the
433 models. For instance, cortical thickness of the superior frontal region consistently predicts cognitive
434 performance (Figure 7), but surface area or volume of the same region is considerably less predictive. In

435 general, the inclusion of additional metrics within a model is associated with lower frequency of a
436 region/tract's survival through regularization, in line with partially overlapping predictive information
437 across different metrics (e.g. Figure 4). In contrast, and somewhat surprisingly, the inclusion of TIV as a
438 covariate had negligible consequences for both the parameter estimates and the overall explanatory
439 power of the model. Focusing on gender as a potential factor by fitting the same model separately yielded
440 similar overall conclusions, apart from an overall higher predictive performance observed in female
441 compared to male. This phenomenon could potentially be explained by variations in motion artefacts
442 (Afacan et al., 2016) or an accelerated cortical maturation process in girls during development
443 (Grabowska, 2017; Fuhrmann et al., 2020, fig 4c).

444

445 Our study has several strengths. The uniquely large childhood sample of ABCD allows us to both explore
446 the optimal model in an exploratory sample of only 15% of the data, and validate it in a test set (Srivastava,
447 2018). As such the sample size enables us to optimize exploration and confirmation, as well as increase
448 parameter precision and power in line with recent recommendations (Marek et al., 2022). Moreover, by
449 using a previously validated measurement model for the latent cognitive variable we further decrease
450 measurement error and increase power and precision (Sauce et al., 2022).

451 However, the main strength of our study is what is commonly absent in the empirical literature: a
452 multimodal analysis including several measures of grey and white matter and the comparison of their
453 explanatory performance. Although previous studies have simultaneously integrated multimodal imaging
454 data (Richard et al., 2020; Rasero et al., 2021), those more commonly examine the individual covariances
455 between multimodal components or factors rather than examine whether distinct metrics provide unique
456 complementary predictive information. The challenge in models with many predictors is that additional
457 predictors will always increase the strength of the joint prediction, at least within sample. In our modeling

458 approach, we rely on three approaches to guard against needless complexity to differentiate the three
459 models: (i) regularization penalizes the predictive estimates downwards to ensure a parsimonious (and
460 often sparse) model; (ii) model comparison criteria which penalize unnecessary complexity, favoring a
461 model with fewer predictors that does 'almost' as well as the more complex model (Jacobucci et al., 2019);
462 and (iii) report effect sizes as adjusted r-squared to account for the number of parameters needed to
463 achieve an overall amount of variance explained.

464

465 Despite these strengths, our study also has several limitations. First and foremost, the present analysis
466 focuses on cross-sectional differences, in a specific age range (9-11 years old). This is a unique
467 developmental period for the brain during which grey matter starts to decrease while white matter
468 continues to increase (Bethlehem et al., 2022), so we expect the precise pattern of brain-behaviour
469 relations to be contingent on this specific developmental period. Future studies will extend our results to
470 a longitudinal approach to tease apart leading and lagging effects and provide additional support in line
471 with causal hypotheses (see King et al., 2018; Vijayakumar et al., 2018). Moreover, although ABCD
472 undertook considerable efforts to recruit a representative sample of the USA, our findings cannot be
473 assumed to generalize to more diverse populations within and especially beyond the US (LeWinn et al.,
474 2017; Garavan et al., 2018).

475 In terms of methodology, the ABCD sample has been collected across different sites and scanners, which
476 may include site or scanner variance we did not incorporate (although Parsons et al., 2023 suggests these
477 site effects in ABCD are considerably smaller than other sources of (un)reliability). Moreover, our findings
478 will be affected by certain pragmatic methodological choices. For example, bilaterally averaging the
479 regions of interest across the hemispheres allowed the possibility to do both regularization and maximum
480 likelihood with a large number of regions. However, it precludes us observing any hemispheric specificity

481 on the prediction of cognitive performance. Finally, the definition of grey and white matter as two
482 different tissues measured by different metrics is inherently an oversimplification. MRI derived measures
483 are proxies of the underlying brain structure and are not able (yet) to isolate specific biophysiological
484 components of grey or white matter. For instance, (apparent) cortical thickness is known to be sensitive
485 to myelinisation of adjacent white matter (Natu et al., 2019), suggesting a partial overlap in statistical
486 contributions of grey and white matter may be as much methodological artefact as biological reality.
487 Finally, segmenting the brain into defined regions reduced structural complexity, correspondingly
488 diminishing explained cognitive performance variance (Fürtjes et al., 2023).

489 To develop a full picture of the complementary roles of grey and white matter to the prediction of
490 cognitive performance, future studies will need to examine large, multimodal, longitudinal and crucially
491 diverse samples across the lifespan. In addition, numerous new measures are needed to estimate cellular
492 mechanisms more accurately. For instance, sulcal depth and curvature analysis provides detailed cortical
493 structure mapping , while myelin water fraction or magnetisation transfer ratio offer more direct
494 approaches to study axons microstructure and myelinisation (Timmers et al., 2016; Lipp et al., 2019).
495 Future work will benefit from metrics that demonstrate a better representation of underlying cellular
496 mechanisms (e.g. Goriounova & Mansvelder, 2019). Beyond the metrics, the emergence of new scanners
497 with higher resolution like the 11.7 Tesla or the CONNECTOM scanners will help unveil the structure within
498 the six layers of the cortex and regional cortico-cortical connectivity (Nowogrodzki, 2018; Huang et al.,
499 2021; Raven et al., 2023).

500

501 The present study was designed to evaluate the extent of the overlap between grey and white matter
502 metrics in the prediction of cognitive performance. Our findings suggest that studies focusing solely on
503 one tissue or one metric when linking brain and cognition are likely missing out on complementary

explanatory power. Studies limited by pragmatic concerns should carefully consider which metric to focus on, informed by the phenotype of interest and the population being studied, and make explicit that the findings are likely contingent on the metrics used in a study. Future work should incorporate a more holistic view of brain structure across modalities, metrics and measures to better elucidate the relationships between brain and cognitive performance.

References:

Afacan, O., Erem, B., Roby, D. P., Roth, N., Roth, A., Prabhu, S. P., & Warfield, S. K. (2016). Evaluation of motion and its effect on brain magnetic resonance image quality in children. *Pediatric Radiology*, *46*(12), 1728–1735. https://doi.org/10.1007/s00247-016-3677-9

Assaf, Y., & Pasternak, O. (2008). Diffusion Tensor Imaging (DTI)-based White Matter Mapping in Brain Research: A Review. *Journal of Molecular Neuroscience*, *34*(1), 51–61. https://doi.org/10.1007/s12031-007-0029-0

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence*, *51*, 10–27. https://doi.org/10.1016/j.intell.2015.04.009

Bethlehem, R. a. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), Article 7906. https://doi.org/10.1038/s41586-022-04554-y

Borgeest, G. S., Henson, R. N., Kietzmann, T. C., Madan, C. R., Fox, T., Malpetti, M., Fuhrmann, D., Knights, E., Carlin, J. D., Cam-CAN, & Kievit, R. A. (2021). *A morphometric double dissociation: Cortical thickness is more related to aging; surface area is more related to cognition* (p. 2021.09.30.462545). bioRxiv. https://doi.org/10.1101/2021.09.30.462545

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Deoni, S. C. L., O'Muircheartaigh, J., Elison, J. T., Walker, L., Doernberg, E., Waskiewicz, N., Dirks, H., Piryatinsky, I., Dean, D. C., & Jumbe, N. L. (2016). White matter maturation profiles through early childhood predict general cognitive ability. *Brain Structure and Function*, *221*(2), 1189–1203. https://doi.org/10.1007/s00429-014-0947-x

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Fjell, A. M., Westlye, L. T., Greve, D. N., Fischl, B., Benner, T., van der Kouwe, A. J. W., Salat, D., Bjørnerud, A., Due-Tønnessen, P., & Walhovd, K. B. (2008). The relationship between diffusion tensor imaging and volumetry as measures of white matter properties. *NeuroImage*, *42*(4), 1654–1668. https://doi.org/10.1016/j.neuroimage.2008.06.005

Fuhrmann, D., Simpson-Kent, I. L., Bathelt, J., Team, T. C., & Kievit, R. A. (2020). A Hierarchical Watershed Model of Fluid Intelligence in Childhood and Adolescence. *Cerebral Cortex (New York, NY)*, *30*(1), 339. https://doi.org/10.1093/cercor/bhz091

Fürtjes, A. E., Cole, J. H., Couvy-Duchesne, B., & Ritchie, S. J. (2023). A quantified comparison of cortical atlases on the basis of trait morphometricity. *Cortex*, *158*, 110–126. https://doi.org/10.1016/j.cortex.2022.11.001

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douaud, G., De Stefano, N., Matthews, P. M., Smith, S. M., Johansen-Berg, H., & James, A. C. (2010). Longitudinal changes in grey and white matter during adolescence. *NeuroImage*, *49*(1), 94–103. https://doi.org/10.1016/j.neuroimage.2009.08.003

Goriounova, N. A., & Mansvelder, H. D. (2019). Genes, Cells and Brain Areas of Intelligence. *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnhum.2019.00044

Grabowska, A. (2017). Sex on the brain: Are gender-dependent structural and functional differences associated with behavior? *Journal of Neuroscience Research*, *95*(1–2), 200–212. https://doi.org/10.1002/jnr.23953

Hagler, D. J., Hatton, SeanN., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B. J., Barch, D. M., Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Bartsch, H., Xue, F., … Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*, *202*, 116091. https://doi.org/10.1016/j.neuroimage.2019.116091

Hoagey, D. A., Rieck, J. R., Rodrigue, K. M., & Kennedy, K. M. (2019). Joint contributions of cortical morphometry and white matter microstructure in healthy brain aging: A partial least squares

correlation analysis. *Human Brain Mapping*, *40*(18), 5315–5329. https://doi.org/10.1002/hbm.24774

Huang, S. Y., Witzel, T., Keil, B., Scholz, A., Davids, M., Dietz, P., Rummert, E., Ramb, R., Kirsch, J. E., Yendiki, A., Fan, Q., Tian, Q., Ramos-Llordén, G., Lee, H.-H., Nummenmaa, A., Bilgic, B., Setsompop, K., Wang, F., Avram, A. V., … Rosen, B. R. (2021). Connectome 2.0: Developing the next-generation ultra-high gradient strength human MRI scanner for bridging studies of the micro-, meso- and macro-connectome. *NeuroImage*, *243*, 118530. https://doi.org/10.1016/j.neuroimage.2021.118530

Hyatt, C. S., Owens, M. M., Crowe, M. L., Carter, N. T., Lynam, D. R., & Miller, J. D. (2020). The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. *NeuroImage*, *205*, 116225. https://doi.org/10.1016/j.neuroimage.2019.116225

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. *Advances in Methods and Practices in Psychological Science*, *2*(1), 55–76. https://doi.org/10.1177/2515245919826527

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *Journal of the American Statistical Association*, *70*(351a), 631–639. https://doi.org/10.1080/01621459.1975.10482485

Kail, R. (1997). The neural noise hypothesis: Evidence from processing speed in adults with multiple sclerosis. *Aging, Neuropsychology, and Cognition*, *4*, 157–165. https://doi.org/10.1080/13825589708256644

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., & Henson, R. N. A. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature Communications*, *5*(1), Article 1. https://doi.org/10.1038/ncomms6658

Kievit, R. A., Lindenberger, U., Goodyer, I. M., Jones, P. B., Fonagy, P., Bullmore, E. T., & Dolan, R. J. (2017). Mutualistic Coupling Between Vocabulary and Reasoning Supports Cognitive Development During Late Adolescence and Early Adulthood. *Psychological Science*, *28*(10), 1419–1431. https://doi.org/10.1177/0956797617710785

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. *Developmental Cognitive Neuroscience*, *33*, 54–72. https://doi.org/10.1016/j.dcn.2017.11.009

Kochunov, P., Glahn, D. C., Lancaster, J., Thompson, P. M., Kochunov, V., Rogers, B., Fox, P., Blangero, J., & Williamson, D. E. (2011). Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan. *NeuroImage*, *58*(1), 41–49. https://doi.org/10.1016/j.neuroimage.2011.05.050

Le Bihan, D. (2003). Looking into the functional architecture of the brain with diffusion MRI. *Nature Reviews Neuroscience*, *4*(6), Article 6. https://doi.org/10.1038/nrn1119

LeWinn, K. Z., Sheridan, M. A., Keyes, K. M., Hamilton, A., & McLaughlin, K. A. (2017). Sample composition alters associations between age and brain structure. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-00908-7

Lipp, I., Jones, D. K., Bells, S., Sgarlata, E., Foster, C., Stickland, R., Davidson, A. E., Tallantyre, E. C., Robertson, N. P., Wise, R. G., & Tomassini, V. (2019). Comparing MRI metrics to quantify white matter microstructural damage in multiple sclerosis. *Human Brain Mapping*, *40*(10), 2917–2932. https://doi.org/10.1002/hbm.24568

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience*, *32*, 67–79. https://doi.org/10.1016/j.dcn.2018.02.006

Magistro, D., Takeuchi, H., Nejad, K. K., Taki, Y., Sekiguchi, A., Nouchi, R., Kotozaki, Y., Nakagawa, S., Miyauchi, C. M., Iizuka, K., Yokoyama, R., Shinada, T., Yamamoto, Y., Hanawa, S., Araki, T., Hashizume, H., Sassa, Y., & Kawashima, R. (2015). The Relationship between Processing Speed and Regional White Matter Volume in Healthy Young People. *PLOS ONE*, *10*(9), e0136386. https://doi.org/10.1371/journal.pone.0136386

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), Article 7902. https://doi.org/10.1038/s41586-022-04492-9

Mueller, R. O., & Hancock, G. R. (2008). Best Practices in Structural Equation Modeling. In J. Osborne, *Best Practices in Quantitative Methods* (pp. 488–508). SAGE Publications, Inc. https://doi.org/10.4135/9781412995627.d38

Muetzel, R. L., Mous, S. E., van der Ende, J., Blanken, L. M. E., van der Lugt, A., Jaddoe, V. W. V., Verhulst, F. C., Tiemeier, H., & White, T. (2015). White matter integrity and cognitive performance in school-age children: A population-based neuroimaging study. *NeuroImage*, *119*, 119–128. https://doi.org/10.1016/j.neuroimage.2015.06.014

Natu, V. S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., Zhen, Z., Cox, S., Weiner, K. S., Weiskopf, N., & Grill-Spector, K. (2019). Apparent thinning of human visual cortex during

childhood is associated with myelination. *Proceedings of the National Academy of Sciences*, *116*(41), 20750–20759. https://doi.org/10.1073/pnas.1904931116

Nowogrodzki, A. (2018). The world's strongest MRI machines are pushing human imaging to new limits. *Nature*, *563*(7729), 24–26. https://doi.org/10.1038/d41586-018-07182-7

Østby, Y., Tamnes, C. K., Fjell, A. M., & Walhovd, K. B. (2011). Morphometry and connectivity of the fronto-parietal verbal working memory network in development. *Neuropsychologia*, *49*(14), 3854–3862. https://doi.org/10.1016/j.neuropsychologia.2011.10.001

Parsons, S., Brandmaier, A. M., Lindenberger, U., & Kievit, R. (2023). *Longitudinal stability of grey matter measures varies across brain regions, imaging metrics, and testing sites in the ABCD study*. PsyArXiv. https://doi.org/10.31234/osf.io/zpu6t

Pulli, E. P., Nolvi, S., Eskola, E., Nordenswan, E., Holmberg, E., Copeland, A., Kumpulainen, V., Silver, E., Merisaari, H., Saunavaara, J., Parkkola, R., Lahdesmaki, T., Saukko, E., Kataja, E.-L., Korja, R., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2023). *Structural brain correlates of non-verbal cognitive ability in 5-year-old children: Findings from the FinnBrain Birth Cohort study* (p. 2023.02.22.529110). bioRxiv. https://doi.org/10.1101/2023.02.22.529110

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W., LaMantia, A.-S., & White, L. (2019). *Neurosciences*. De Boeck Superieur.

Rasero, J., Sentis, A. I., Yeh, F.-C., & Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLOS Computational Biology*, *17*(3), e1008347. https://doi.org/10.1371/journal.pcbi.1008347

Raven, E. P., Veraart, J., Kievit, R. A., Genc, S., Ward, I. L., Hall, J., Cunningham, A., Doherty, J., van den Bree, M. B. M., & Jones, D. K. (2023). In vivo evidence of microstructural hypo-connectivity of brain white matter in 22q11.2 deletion syndrome. *Molecular Psychiatry*, 1–11. https://doi.org/10.1038/s41380-023-02178-w

Richard, H., Gresele, L., Hyvarinen, A., Thirion, B., Gramfort, A., & Ablin, P. (2020). Modeling Shared responses in Neuroimaging Studies through MultiView ICA. *Advances in Neural Information Processing Systems*, *33*, 19149–19162. https://proceedings.neurips.cc/paper/2020/hash/de03beffeed9da5f3639a621bcab5dd4-Abstract.html

Ritchie, S. J., Booth, T., Valdés Hernández, M. del C., Corley, J., Maniega, S. M., Gow, A. J., Royle, N. A., Pattie, A., Karama, S., Starr, J. M., Bastin, M. E., Wardlaw, J. M., & Deary, I. J. (2015). Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence*, *51*, 47–56. https://doi.org/10.1016/j.intell.2015.05.001

Rolls, E. T., & Deco, G. (2015). Stochastic cortical neurodynamics underlying the memory and cognitive changes in aging. *Neurobiology of Learning and Memory*, *118*, 150–161. https://doi.org/10.1016/j.nlm.2014.12.003

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Sauce, B., Liebherr, M., Judd, N., & Klingberg, T. (2022). The impact of digital media on children's intelligence while controlling for genetic differences in cognition and socioeconomic background. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-11341-2

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). *Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures*. *8*(2), 52.

Schnack, H. G., van Haren, N. E. M., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., Kahn, R. S., & Hulshoff Pol, H. E. (2015). Changes in Thickness and Surface Area of the Human Cortex and Their Relationship with Intelligence. *Cerebral Cortex*, *25*(6), 1608–1617. https://doi.org/10.1093/cercor/bht357

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Srivastava, S. (2018). *Sound Inference in Complicated Research: A Multi-Strategy Approach*. PsyArXiv. https://doi.org/10.31234/osf.io/bwr48

Tamnes, C. K., Fjell, A. M., Westlye, L. T., Østby, Y., & Walhovd, K. B. (2012). Becoming Consistent: Developmental Reductions in Intraindividual Variability in Reaction Time Are Related to White Matter Integrity. *Journal of Neuroscience*, *32*(3), 972–982. https://doi.org/10.1523/JNEUROSCI.4779-11.2012

Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2010). Brain Maturation in Adolescence and Young Adulthood: Regional Age-Related Changes in Cortical Thickness and White Matter Volume and Microstructure. *Cerebral Cortex*, *20*(3), 534–548. https://doi.org/10.1093/cercor/bhp118

Tamnes, C. K., Østby, Y., Walhovd, K. B., Westlye, L. T., Due-Tønnessen, P., & Fjell, A. M. (2010). Neuroanatomical correlates of executive functions in children and adolescents: A magnetic resonance imaging (MRI) study of cortical thickness. *Neuropsychologia*, *48*(9), 2496–2508. https://doi.org/10.1016/j.neuropsychologia.2010.04.024

Timmers, I., Roebroeck, A., Bastiani, M., Jansma, B., Rubio-Gozalbo, M., & Zhang, H. (2016). Assessing Microstructural Substrates of White Matter Abnormalities: A Comparative Study Using DTI and NODDI. *PLoS ONE*, *11*, e0167884. https://doi.org/10.1371/journal.pone.0167884

Vijayakumar, N., Mills, K. L., Alexander-Bloch, A., Tamnes, C. K., & Whittle, S. (2018). Structural brain development: A review of methodological approaches and best practices. *Developmental Cognitive Neuroscience*, *33*, 129–148. https://doi.org/10.1016/j.dcn.2017.11.008

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., Conway, K., Deeds, B. G., Dowling, G. J., Grant, S., Howlett, K. D., Matochik, J. A., Morgan, G. D., Murray, M. M., Noronha, A., Spong, C. Y., … Weiss, S. R. B. (2018).

The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, *32*, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Wandell, B. A. (2016). Clarifying Human White Matter. *Annual Review of Neuroscience*, *39*(1), 103–128. https://doi.org/10.1146/annurev-neuro-070815-013815

Xing, Y., Yang, J., Zhou, A., Wang, F., Wei, C., Tang, Y., & Jia, J. (2021). White Matter Fractional Anisotropy Is a Superior Predictor for Cognitive Impairment Than Brain Volumes in Older Adults With Confluent White Matter Hyperintensities. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.633811

Zhao, Y., Li, Q., Du, J., He, H., Liang, P., Lu, J., & Li, K. (2021). Exploring the Relationship between Gray and White Matter in Healthy Adults: A Hybrid Research of Cortical Reconstruction and Tractography. *BioMed Research International*, *2021*, 6628506. https://doi.org/10.1155/2021/6628506