Title The (Limited?) Utility of Brain Age as a Biomarker for Capturing Cognitive Decline

Alina Tetereva[a] https://orcid.org/0000-0003-2077-628X

Narun Pat[a] https://orcid.org/0000-0003-1459-5255

[a]Department of Psychology, University of Otago, New Zealand, 9016

Corresponding Author:
Narun Pat, PhD, also known as Narun Pornpattananangkul
Department of Psychology, University of Otago
William James Building
275 Leith Walk
Dunedin 9016, New Zealand
Email: narun.pat@otago.ac.nz

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

Acknowledgements
Data were provided by the Human Connectome Project in Aging. Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number U01AG052564 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The author(s) wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL https://www.nesi.org.nz. A.T. and N.P. were supported by Health Research Council Funding (21/618) and by the University of Otago.

Abstract

For decades, neuroscientists have been on a quest to search for a biomarker that can help capture age-related cognitive decline. One well-known candidate is Brain Age, or a predicted value based on machine-learning models built to predict chronological age from brain MRI data. Here we aim to formally evaluate the utility of Brain Age as a biomarker for capturing cognitive decline. Using 504 aging participants (36-100 years old) from the Human Connectome Project in Aging, we created 26 age-prediction models for Brain Age based on different combinations of MRI modalities. We first tested how much Brain Age from these age-prediction models added to what we had already known from a person's chronological age in capturing cognitive decline. Based on the commonality analyses, we found a large degree of overlap between Brain Age and chronological age, so much so that, at best, Brain Age could uniquely add only around 1.6% in explaining variation in cognitive decline. Next, the age-prediction models that performed better at predicting chronological age did NOT necessarily create better Brain Age for capturing cognitive decline over and above chronological age. Instead, better-performing age-prediction models created Brain Age that overlapped larger with chronological age, up to around 29% out of 32%, in explaining cognitive decline, thus not improving the models' ability to capture cognitive decline. Lastly, unlike Brain Age, Brain Cognition, or a predicted value based on machine-learning models built to predict cognitive abilities from brain MRI data, provided much higher unique effects. Brain Cognition added over 11% to explain variation in cognitive decline beyond chronological age, leading to around a 1/3-time improvement of the total variation explained. Accordingly, while demonstrating the limited utility of Brain Age, we provided a solution to improve our ability to use brain MRI data as a biomarker for cognitive decline.

Keywords: Brain Age, Cognitive Decline, brain MRI, machine learning, Aging, biomarker

Introduction

Older adults often experience declines in several cognitive abilities such as memory, attention and processing speed, collectively known as fluid cognition ($cognition_{fluid}$) (Salthouse, 2019; Weintraub et al., 2014). Having objective biomarkers to capture these declines in $cognition_{fluid}$ would give researchers and clinicians a tool to detect early cognitive impairments, monitor treatment/intervention efficacy and forecast cognitive prognosis (Frisoni et al., 2017). Over the past decade, Brain Age (Franke et al., 2010) has emerged as a potential biomarker to capture declines in $cognition_{fluid}$ (Cole, 2020; Cole et al., 2018; Liem et al., 2017; Richard et al., 2018; Wrigglesworth et al., 2022; see review Boyle et al., 2021). Yet, to justify the use of Brain Age as an informative biomarker for $cognition_{fluid}$, we still need to address many outstanding issues.

First, to what extent does having information on Brain Age improve our ability to capture declines in $cognition_{fluid}$ beyond knowing a person's chronological age? To compute Brain Age, researchers first built a prediction model that predicts the chronological age based on a person's brain neuroimaging data (Baecker et al., 2021). They then apply this prediction model to an unseen individual, not part of the model-building process. Brain Age is the predicted value of this model. Accordingly, by design, Brain Age is tightly close to chronological age. Because chronological age usually has a strong relationship with $cognition_{fluid,}$ to begin with, it is unclear how much Brain Age adds to what is already captured by chronological age.

Note researchers often subtract chronological age from Brain Age, creating an index known as Brain Age Gap (Franke & Gaser, 2019). A higher value of Brain Age Gap is thought to reflect accelerated/premature aging. Yet, given that Brain Age Gap is calculated based on both Brain Age and chronological age, Brain Age Gap still depends on chronological age (Butler et al., 2021). If, for instance, Brain Age was based on prediction models with poor performance and made a prediction that everyone was 50 years old, individual differences in Brain Age Gap would then depend solely on chronological age (i.e., 50 minus chronological age). In addition to the dependency on chronological age, Brain Age is often overestimated in younger, but underestimated in older, individuals (Le et al., 2018). There are many adjustments proposed to correct for this estimation bias, but the outcomes tend to be similar, if not identical, to each other (Beheshti et al., 2019; de Lange & Cole, 2020; Liang et al., 2019; Smith et al., 2019). These adjustments can be applied to Brain Age and Brain Age Gap, creating Corrected Brain Age and Corrected Brain Age Gap, respectively. Corrected Brain Age Gap in particular is viewed as being able to control for both age dependency and estimation biases (Butler et al., 2021). Here we tested the ability of different Brain Age calculations in capturing $cognition_{fluid}$, over and above chronological age.

Second, do better-performing age-prediction models correspond to the improvement in an ability to capture $cognition_{fluid}$? Over the past decades, there has been a race to improve the performance of age-prediction models to be better at predicting chronological age, for instance, by combining different MRI/neuroimaging modalities features (Cole, 2020; Engemann et al., 2020; Liem et al., 2017) or by applying more sophisticated machine-learning algorithms (Baecker et al., 2021; Jonsson et al., 2019; Zhao & Zhao, 2021). However, the improvement in predicting chronological age may not necessarily make Brain

Age to be better at capturing $cognition_{fluid}$. If, for instance, the age-prediction model had the perfect performance, Brain Age Gap would be exactly zero and would have no utility in capturing $cognition_{fluid}$ beyond chronological age. Few, if any, studies have examined the performance of age-prediction models as a function of their ability to capture $cognition_{fluid}$.

Third and finally, can we further improve our ability to capture the decline in $cognition_{fluid}$ by using, not only Brain Age and chronological age, but also another biomarker, Brain Cognition? Analogous to Brain Age, Brain Cognition is defined as a predicted value from prediction models that predict $cognition_{fluid}$ based on a person's brain data (Dubois et al., 2018; Pat, Wang, Anney, et al., 2022; Rasero et al., 2021; Sripada et al., 2020; Tetereva et al., 2022; for review, see Vieira et al., 2022). Age-related cognitive decline is not only related to the changes in age, but also to the changes in cognition. While Brain Age is about the former, Brain Cognition is about the letter. Brain Age and Brain Cognition are found to capture some overlapped variation in cognitive abilities (Jiang et al., 2022). Despite the overlap, there might still be unique effects contributed by each biomarker that, when combined, may collectively help capture $cognition_{fluid}$.

Our study set out to test the utility of Brain Age as a biomarker for capturing decline in $cognition_{fluid}$. Using aging participants (36-100 years old) from the Human Connectome Project in Aging, we computed different Brain Age indices (including Brain Age, Brain Age Gap, Corrected Brain Age and Corrected Brain Age Gap) and Brain Cognition from prediction models based on different sets of MRI features. These MRI features covered task, resting-state and structural MRI, creating 26 prediction models in total. We, then, tested the biomarkers' ability to explain $cognition_{fluid}$ in unseen participants. To test the ability of Brain Age by itself, we applied simple regression models with each Brain Age index as a sole regressor to explain $Cognition_{fluid}$. Next, to test the unique effects of Brain Age in explaining $Cognition_{fluid}$ beyond chronological age, we applied multiple regression models with both each Brain Age index and chronological age as regressors to explain $Cognition_{fluid}$. To reveal how much chronological age and Brain Age had in common in explaining $Cognition_{fluid}$ (i.e., common effects), we then applied the commonality analysis (Nimon et al., 2008) to these multiple regression models. Additionally, given that certain sets of MRI features led to prediction models that were better at predicting chronological age, we also examined if these better-performing age-prediction models improved the ability of Brain Age in explaining $cognition_{fluid}$. Finally, to test whether Brain Cognition could further improve the ability to capture $cognition_{fluid}$, beyond Brain Age and chronological age, we added Brain Cognition to the multiple regression models and tested its unique and common effects in explaining $Cognition_{fluid}$.

Results

Relationship between chronological age and $Cognition_{fluid}$

Figure 1a shows the negative relationship between chronological age and $Cognition_{fluid}$ ($r(502) = -.57$, $p < .001$, $R^2 = .32$). Older individuals tended to have a lower $Cognition_{fluid}$ score.

Predictive performance of prediction models for Brain Age and Brain Cognition

Figure 1b and 1c show the predictive performance of different sets of brain MRI features in predicting chronological age and $\text{Cognition}_{\text{fluid}}$, respectively. For age prediction, the top-four similarly performing models were 'stacked' models that included multiple sets of brain MRI features: "Stacked: All excluding Task Contrast", "Non Task", "All excluding Task FC" and "All" ($R^2 > .76$, $r > .83$, MAE < 65 months). For $\text{Cognition}_{\text{fluid}}$ prediction, the top-performing model was "Stacked: All" ($R^2 = .393$, $r = .627$, MAE = 7.7 points). The best set of features across age and $\text{Cognition}_{\text{fluid}}$ prediction was cortical thickness. Across sets of MRI features, the age-prediction models tended to provide higher R^2 and r than the $\text{Cognition}_{\text{fluid}}$-prediction models. Figure 2 shows the feature importance of prediction models based on each of the 18 sets of features. Figure 3 shows the feature importance of the eight stacked prediction models.

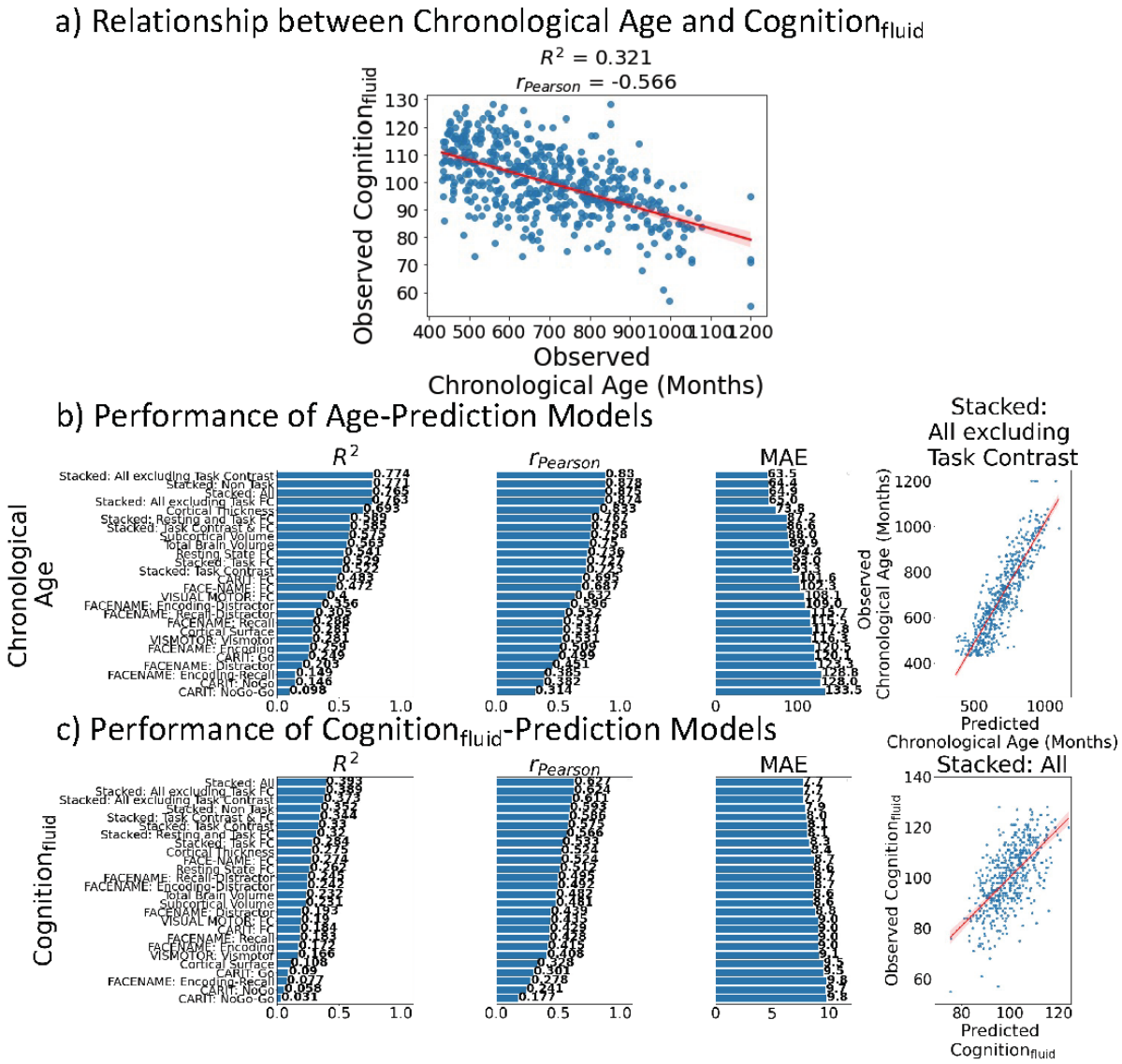


Figure 1. Relationship between chronological age and $\text{Cognition}_{\text{fluid}}$ (a) and predictive performance of prediction models using Brain MRI from different sets of MRI features to predict chronological age (b) and $\text{Cognition}_{\text{fluid}}$ (c). Note we only provided the scatter plots between observed and predicted values in the test sets from the best prediction models for each target here. See Supplementary Figures 1 and 2 for the scatter plots from other prediction models.

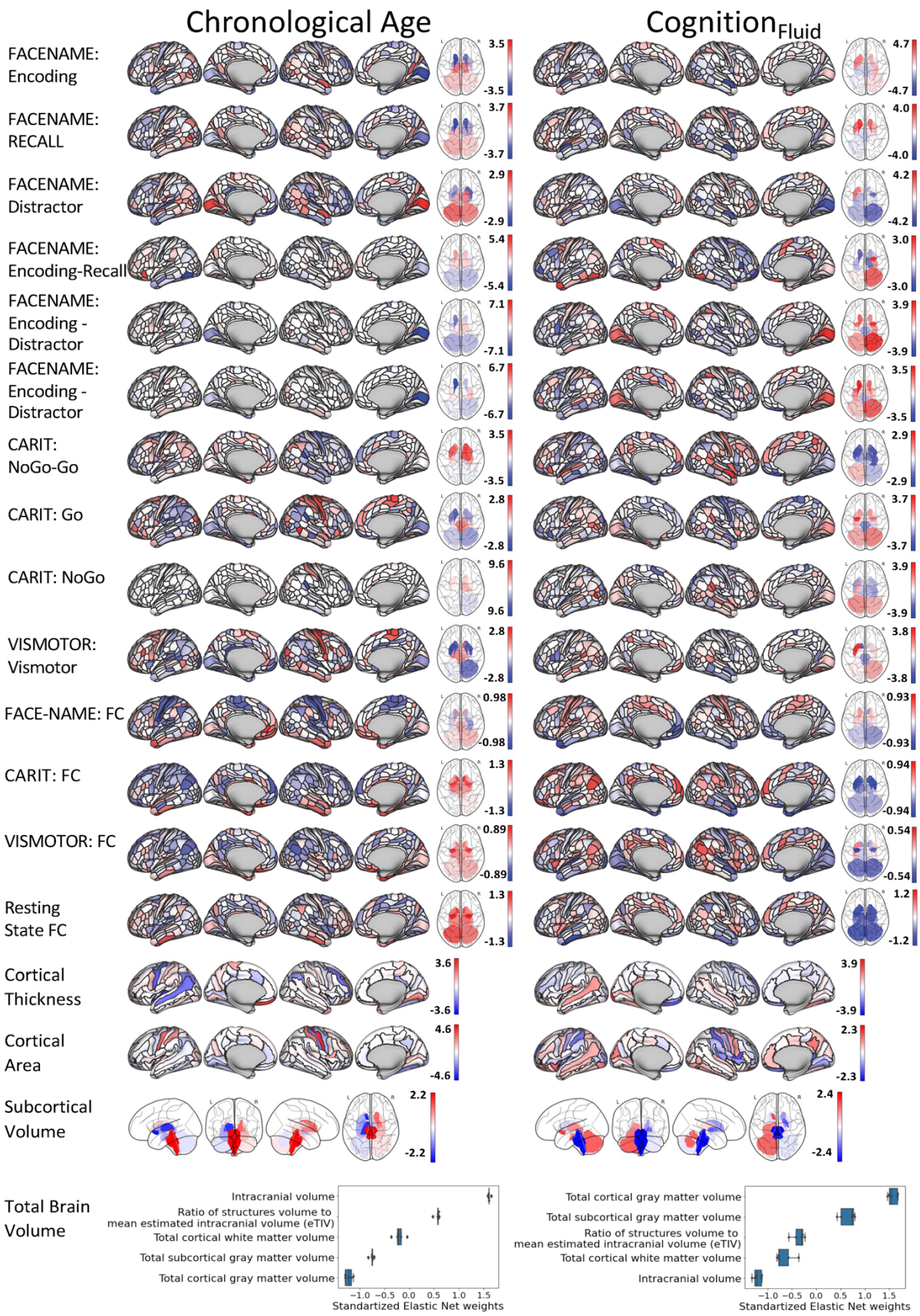


Figure 2. Feature importance of prediction models based on each of the 18 sets of features. We calculated feature importance by, first, standardising Elastic Net weights across brain features of each set of features from each test fold. We then plotted the averaged weights

across the five test folds for each of the set of features. For functional connectivity (FC), we, first, multiplied the absolute PCA scores (extracted from the 'components_' attribute of 'sklearn.decomposition.PCA') with Elastic Net weights and, then, summed the multiplied values across the 75 components, leaving 71,631 ROI-pair indices. Thereafter, we standardised these indices from each test fold and averaged them across the five test folds. Finally, given that the 71,631 ROI-pair indices were based on correlations among 379 ROIs, we averaged the ROI-pair indices from each ROI and plotted them. Accordingly, our FC plots showed the contribution of each seeding area.

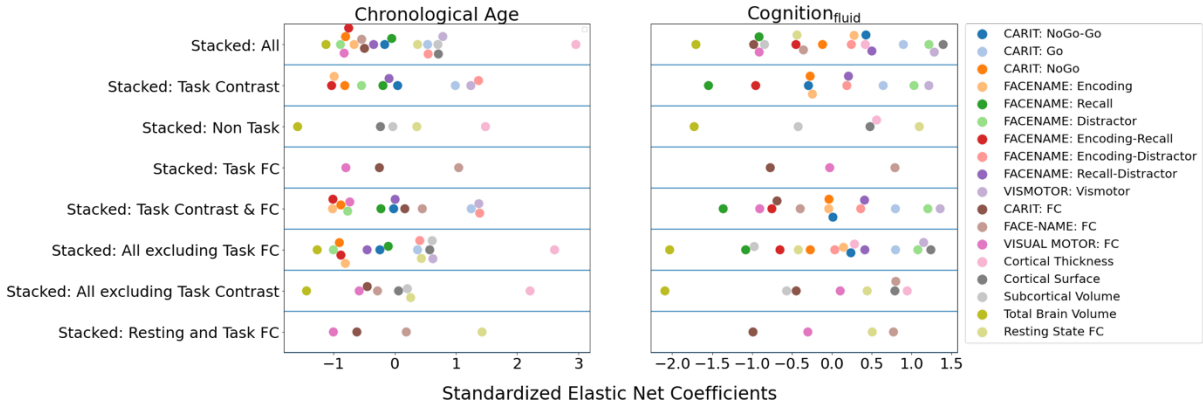
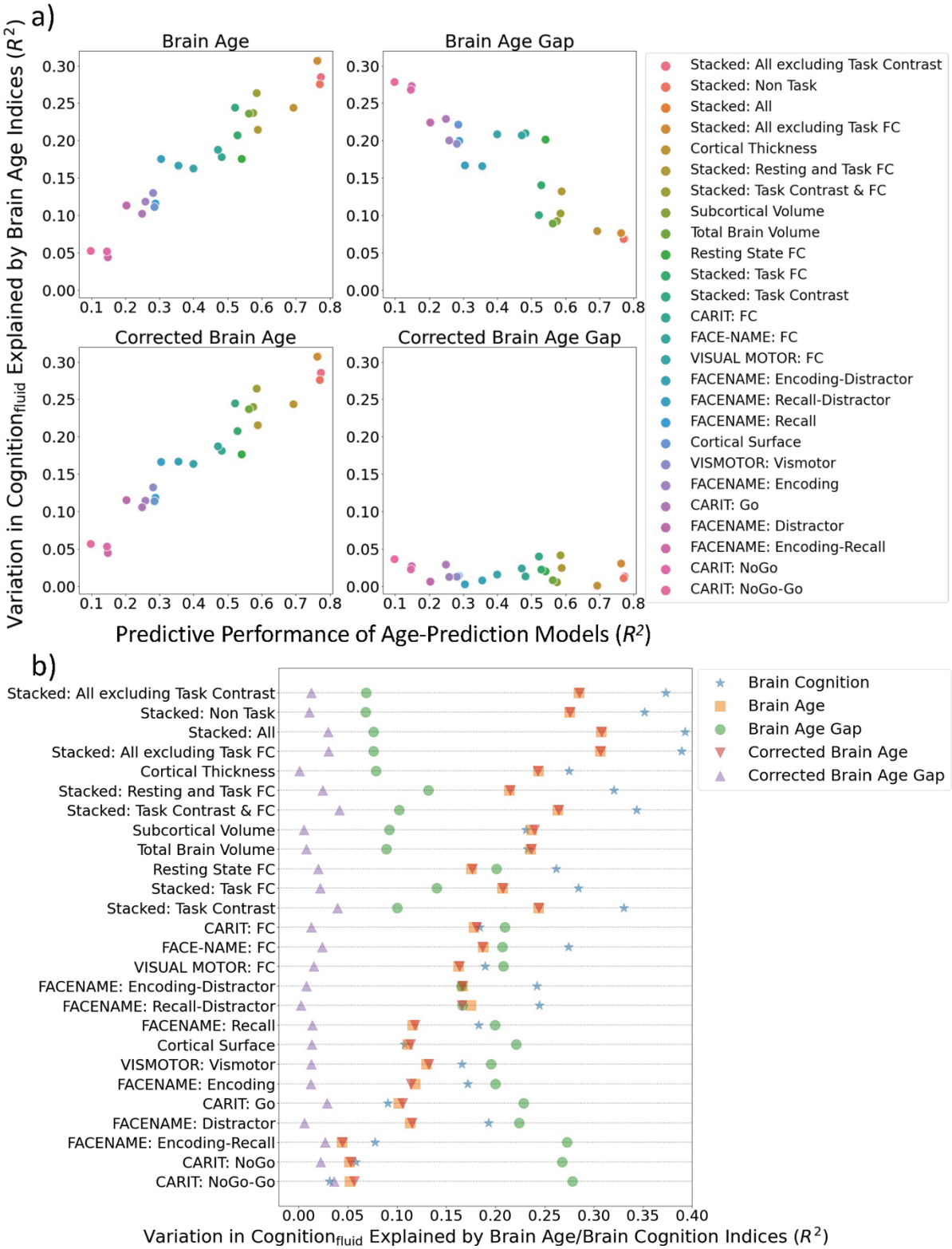


Figure 3. Feature importance of the eight stacked prediction models. Here we plotted the Elastic weights, standardised across predicted values from different sets of features and averaged across the five test folds.

Simple regression: Using each Brain Age index to explain Cognition$_{fluid}$

Figure 4a shows variation in Cognition$_{fluid}$ explained by Brain Age Indices when having each Brain Age index as the sole regressor in simple regression models. Brain Age and Corrected Brain Age created from higher-performing age-prediction models explained a higher amount of variation in Cognition$_{fluid}$. However, Brain Age Gap created from the *lower*-performing age-prediction models explained a higher amount of variation in Cognition$_{fluid}$. For instance, the top performing age-prediction model, "Stacked: All excluding Task Contrast", generated Brain Age and Corrected Brain Age that explained the highest amount of variation in Cognition$_{fluid}$, but, at the same time, produced Brain Age Gap that explained the least amount of variation in Cognition$_{fluid}$.

On the contrary, an amount of variation in Cognition$_{fluid}$ explained by Corrected Brain Age Gap was relatively small (maximum at R^2=.041) across age-prediction models and did not relate to the predictive performance of the age-prediction models. Figure 4b shows variation in Cognition$_{fluid}$ explained by Brain Cognition, as compared to Brain Age indices. Brain Cognition appeared to explain a higher amount of variation in Cognition$_{fluid}$ than any Brain Age indices, especially for top-performing age/cognition-prediction models (e.g., Stacked: All).



<u>Figure 4. Simple regression: using each Brain Age index or Brain Cognition to explain Cognition$_{fluid}$</u>. 4a shows variation in Cognition$_{fluid}$ explained by each Brain Age index as a function of the predictive performance of age-prediction models. 4b plots variation in Cognition$_{fluid}$ explained by Brain Age indices and Brain Cognition.

Multiple regression: Using chronological age and each Brain Age index to explain $\text{Cognition}_{\text{fluid}}$

Figure 5 shows the commonality analysis of multiple regression models, having both chronological age and each Brain Age index as the regressors for $\text{Cognition}_{\text{fluid}}$. We found R^2 for these models at M = .326 (SD = 005). The unique effects of Brain Age indices were all relatively small (maximum at $\Delta R^2_{\text{Brain Age index}}$ = .0161, with statistically significant at *p-value* < .05 in 10 out of 26 models) across the four Brain Age indices and across different age-prediction models.

However, it is clear that different Brain Age indices led to different levels of the unique effects of chronological age and the common effects between chronological age and Brain Age indices. For the top-performing age-prediction models (e.g., Stacked: All excluding Task Contrast), the unique effects of chronological age were low for Brain Age and Corrected Brain Age, but high for Brain Age Gap. On the contrary, the lower-performing age-prediction models provided high common effects for Brain Age and Corrected Brain Age, but low for Brain Age Gap. Nonetheless, for Corrected Brain Age Gap, the unique effects of chronological age were much higher than the common effects across all age-prediction models.

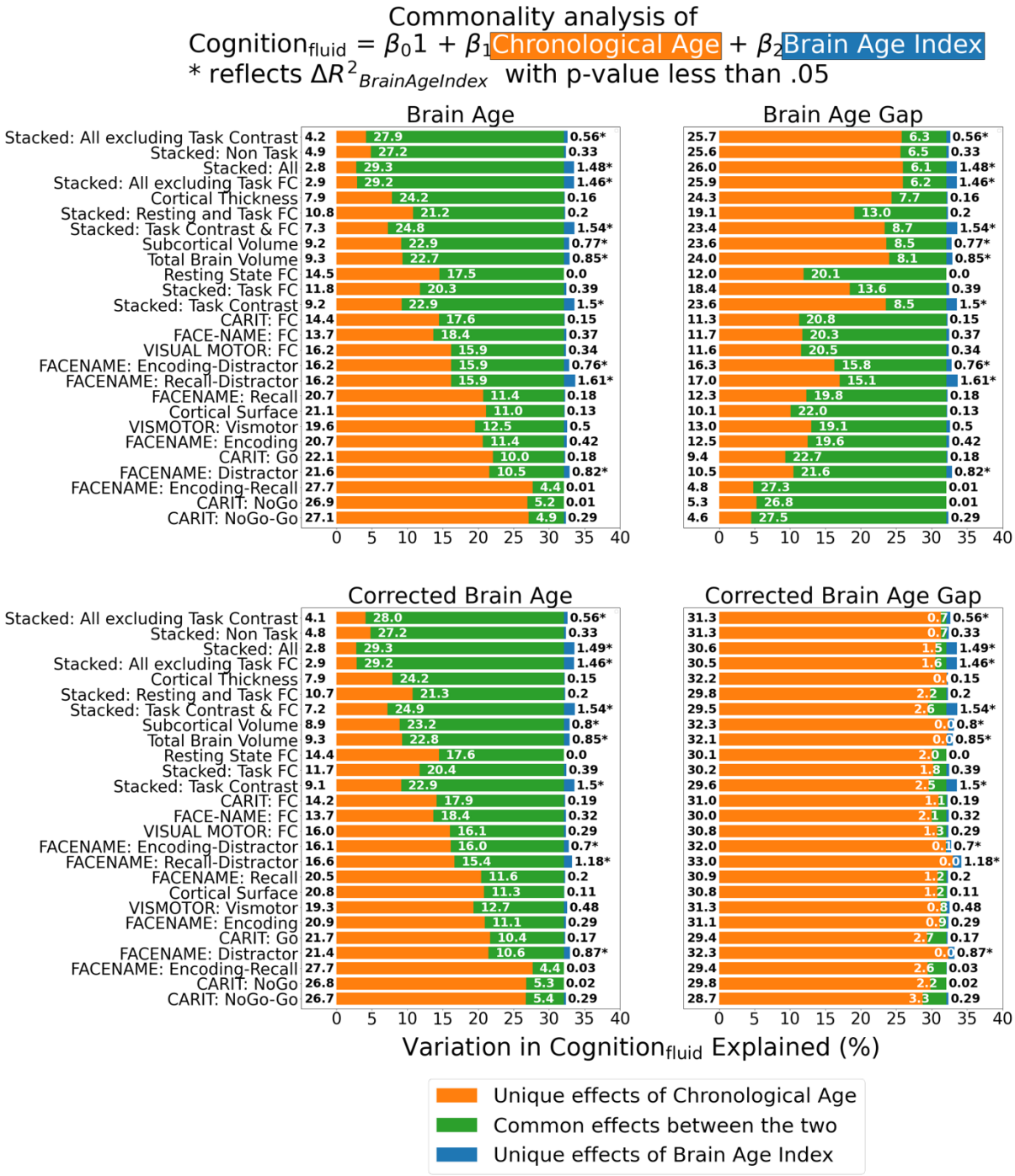


Figure 5. Commonality analysis of a multiple regression model, having both chronological age and each Brain Age index as the regressors for Cognition$_{fluid}$.

Multiple regression: Using chronological age, each Brain Age index and Brain Cognition to explain Cognition$_{fluid}$

Figure 6 shows the commonality analysis of multiple regression models, having chronological age, each Brain Age index and Brain Cognition as the regressors for Cognition$_{fluid}$. We found R^2 for these models at M=.385 (SD=.042). As before, the unique effects of Brain Age indices were all relatively small across the four Brain Age indices and across different prediction models. On the contrary, the unique effects of Brain Cognition

appeared much larger (maximum at $\Delta R^2_{cognition}$ = .1183, statistically significant *p-value* at .05 in 24 out of 26 models).

For top-performing age/cognition-prediction models (e.g., Stacked All), the largest proportion of $Cognition_{fluid}$ was attributed to a) the common effects among the three for Brain Age and Corrected Brain Age and b) the common effects between chronological age and Brain Cognition for Brain Age Gap and Corrected Brain Age Gap.

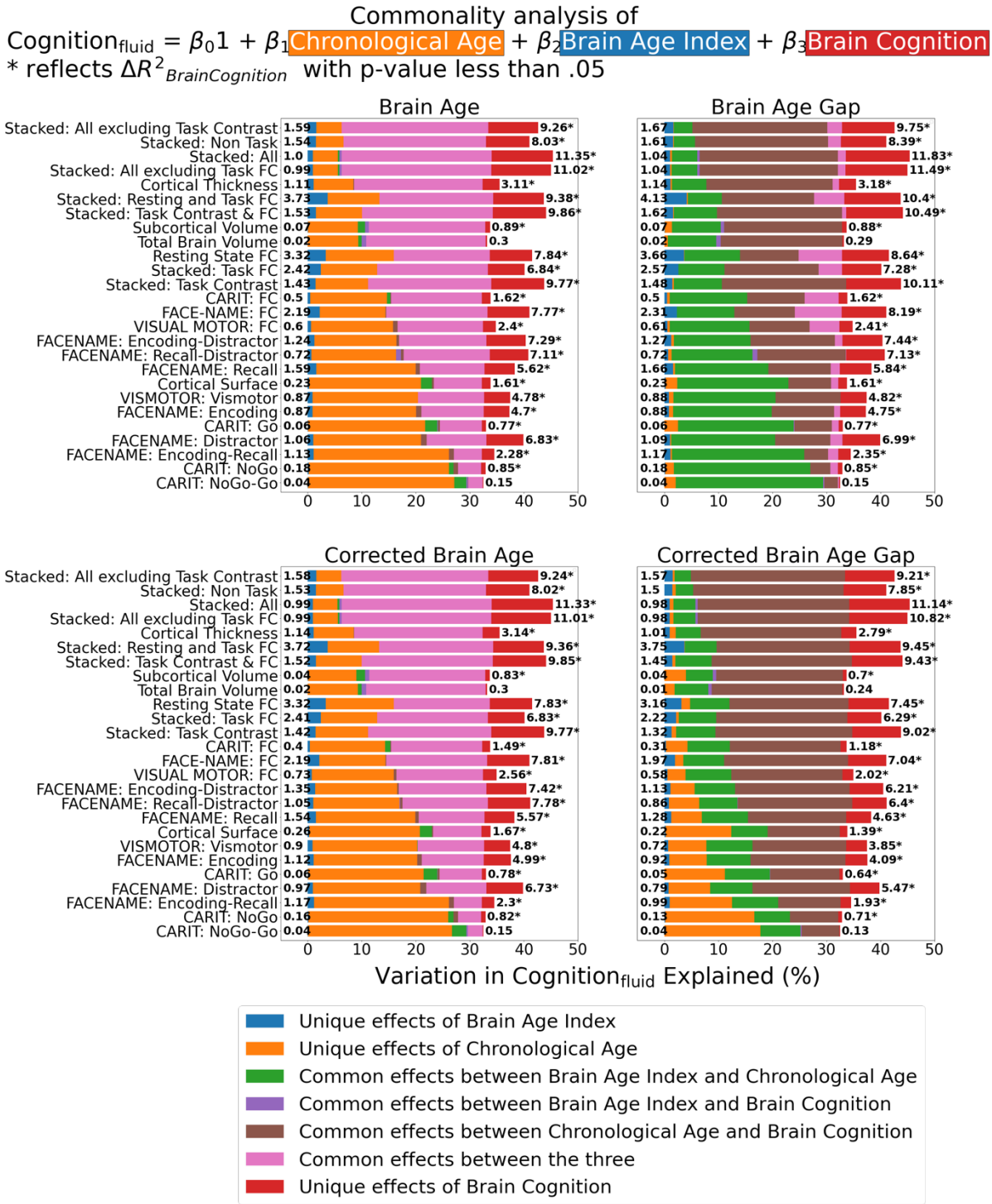


Figure 6. Commonality analysis of a multiple regression model, having chronological age and each Brain Age index and Brain Cognition as the regressors for $Cognition_{fluid}$.

Discussion

To demonstrate the utility of Brain Age as a biomarker for $cognition_{fluid}$, we investigated three unsettling issues. First, how much does Brain Age add to what is already captured by chronological age? The short answer is very little. Second, do better-performing age-prediction models improve the ability of Brain Age to capture $cognition_{fluid}$? Unfortunately, the answer is no. Third, do we have a solution that can improve our ability to capture $cognition_{fluid}$ from brain MRI? The answer is, fortunately, yes. Using Brain Cognition as a biomarker, along with chronological age, seemed to capture a higher amount of variation in $cognition_{fluid}$ than only using Brain Age.

First, Brain Age itself did not add much more information to help us capture $cognition_{fluid}$ than what we had already known from a person's chronological age. This can clearly be seen from the small unique effects of Brain Age indices in the multiple regression models having Brain Age and chronological age as the regressors. While the unique effects of some Brain Age indices from certain age-prediction models were statistically significant, there were all relatively small. Without Brain Age indices, chronological age by itself already explained around 32% of variation in $cognition_{fluid}$. Including Brain Age indices only added around 1.6% at best.

Investigating the simple regression models and the commonality analysis between each Brain Age index and chronological age provided additional insights. In the simple regression models, higher-performing age-prediction models, such as stacked models, created Brain Age and Corrected Brain Age that captured a higher amount of variation in $Cognition_{fluid}$. Because both Brain Age and Corrected Brain Age from higher-performing age-prediction models were closer to the real chronological age of participants, their ability to capture $cognition_{fluid}$ mirrored the ability of chronological age. The commonality analysis confirmed this by showing higher common effects between (Corrected) Brain Age and chronological age from higher-performing age-prediction models. In contrast, lower-performing (as opposed to higher-performing) age-prediction models, such as CARIT NoGo-Go, created Brain Age Gap that explained a higher amount of variation in $Cognition_{fluid}$. Brain Age Gap was a result of subtracting a real chronological age from Brain Age. And when Brain Age was a poor indicator of the real chronological age, the ability of Brain Age Gap is driven more by the real chronological age (Butler et al., 2021). The commonality analysis confirmed this by showing higher common effects, therefore more similarity in variance, between Brain Age Gap and chronological age from lower-performing, than higher-performing, age-prediction models.

Corrected Brain Age Gap, on the other hand, showed weak effects in the simple regression models across all age-prediction models (max at around 4.1% of variation explained). Corrected Brain Age Gap was the only index among the four that appeared to deconfound the influences of chronological age on the relationship between brain aging and $cognition_{fluid}$ (Butler et al., 2021). This can be seen in the small common effects between Corrected Brain Age Gap and chronological age in the multiple regression models with chronological age and each Brain Age index as regressors. Note while these common effects between Corrected Brain Age Gap and chronological age were small, most were not zero (max at around 3.3% of variation explained). This means that the correction done to

deconfound the influences of chronological age on Corrected Brain Age Gap (de Lange & Cole, 2020) may not be perfect. Perhaps this is because the estimation of the influences of chronological age was done in the training set, which might not fully be applicable to the test sets. Still, weak effects of Corrected Brain Age Gap in the simple regression indicate that, after controlling for the influences of chronological age, this Brain Age index could only account for a small amount of variation in $\text{cognition}_{\text{fluid}}$. In other words, the weak effects of Corrected Brain Age Gap shown by the simple regression are consistent with the small unique effects across the four Brain Age indices shown by the multiple regression models having a Brain Age index and chronological age as regressors.

Second, the predictive performance of age-prediction models did not correspond to the ability of Brain Age to capture $\text{cognition}_{\text{fluid}}$ over and above chronological age. For instance, while the best-performing age-prediction model was "Stacked: All excluding Task Contrast" (R^2=.775), the unique effects of Brain Age indices from this model in the two-regressor multiple regressions were weak ($\Delta R^2_{\text{Brain Age index}} \leq .0048$) and not statistically significant. The highest unique effects of Brain Age indices in the two-regressor multiple regression models were from the FACENAME: Distractor model ($\Delta R^2_{\text{Brain Age index}} \leq .0135$, $p < .05$) that had a poorer performance in predicting chronological age (R^2=.204). Accordingly, a race to improve the performance of age-prediction models (Baecker et al., 2021) does not necessarily enhance the utility of Brain Age indices as a biomarker for $\text{cognition}_{\text{fluid}}$. This calls for a new paradigm. Future research should aim to build prediction models for Brain Age indices that are not necessarily good at predicting age, but at capturing phenotypes of interest, such as $\text{cognition}_{\text{fluid}}$ and beyond.

Third, while Brain Age indices may not add much to capture $\text{cognition}_{\text{fluid,}}$ over and above chronological age, Brain Cognition did as shown by its unique effects. Adding Brain Cognition as a regressor along with chronological age and a Brain Age index allowed us to explain as much as 11% more of the variation in $\text{cognition}_{\text{fluid}}$ (i.e., around 1/3 times better than the multiple regression models without Brain Cognition). Based on the commonality analyses, the multiple regression models that explained higher variation were driven predominantly by the common effects involving Brain Cognition. This overlapped variation was consistent with a recent study (Jiang et al., 2022). Altogether, while both the effects of Brain Age and Brain Cognition overlapped with chronological age, Brain Cognition offered additional unique effects that helped improve the ability to capture $\text{cognition}_{\text{fluid}}$.

What does it mean then for researchers/clinicians who would like to use Brain Age as a biomarker? First, they have to be aware of the overlap in variation between Brain Age and chronological age and should focus on the contribution of Brain Age over and above chronological age. Using Brain Age Gap will not fix this. Similar to a previous recommendation (Butler et al., 2021), we suggest focusing on Corrected Brain Age Gap or, better, unique effects of Brain Age indices in multiple regressions. In the case of $\text{cognition}_{\text{fluid}}$, the unique effects might be too small to be clinically meaningful. While our emphasis was on $\text{cognition}_{\text{fluid}}$, we suspect that the unique effects of Brain Age might also be small for other phenotypes, especially if those phenotypes vary with age (e.g., neurodegenerative risks). Next, researchers/clinicians should not select the MRI modalities or machine-learning algorithms based on age-prediction performance, but rather select them based on the unique effects of Brain Age indices. Finally, beyond Brain Age,

researchers/clinicians should explore the unique effects of other biomarkers that may tap closer into the phenotypes of interest. One straightforward approach tested here is to build the model using the phenotype of interest as a target (i.e., $cognition_{fluid}$). This approach significantly improved our ability to capture $cognition_{fluid}$ in the current study.

Methods and Materials

Dataset

We used the Human Connectome Project in Aging (HCP-A) (Bookheimer et al., 2019) Release 2.0 (24-February-2021). HCP-A's 'typical-aging' participants (36-100 years old) may have prevalent health conditions (e.g., hypertension and different forms of vascular risks) but did not have identified pathological causes of cognitive decline (e.g., stroke and clinical dementia). In this Release, HCP-A provided data from 725 participants. HCP-A offered quality control flags, and here, we removed participants with the flag 'A' anatomical anomalies or 'B' segmentation and surface (n= 117). Following further removal of participants with missing values in any of MRI modalities (n=15) or cognitive measurements (n= 111), we ultimately included 504 individuals (293 females, M= 57.83 (SD=14.25) years old) in our analyses. For ethical procedures including informed consent, please see Bookheimer and colleagues' (2019).

Sets of brain MRI features

HCP-A provides details of parameters for brain MRI elsewhere (Bookheimer et al., 2019; Harms et al., 2018). Here we used MRI data that were pre-processed by the HCP-A with recommended methods, including the MSMALL alignment (Glasser et al., 2016; Robinson et al., 2018) and ICA-FIX (Glasser et al., 2016) for functional MRI. We used multiple brain MRI modalities, covering task functional MRI (task fMRI), resting-state functional MRI (rsfMRI) and structural MRI (sMRI), and organised them into 19 sets of features.

<u>Sets of Features 1-10: Task fMRI contrast (Task Contrast)</u>

Task contrasts reflect fMRI activation relevant to events in each task. Bookheimer and colleagues (2019) provided detailed information about the fMRI in HCP-A. Here we focused on the pre-processed task fMRI files with a suffix, "_PA_Atlas_MSMAll_hp0_clean.dtseries.nii." To extract Task Contrasts, we regressed the fMRI time series on the convolved task events using a double-gamma canonical hemodynamic response function via FMRIB Software Library (FSL)'s FMRI Expert Analysis Tool (FEAT) (Woolrich et al., 2001). We then parcellated the contrast 'cope' files, using the Glasser atlas (Gordon et al., 2016) for cortical surface regions and the Freesurfer's automatic segmentation (aseg) (Fischl et al., 2002) for subcortical regions. This resulted in 379 regions, which was, in turn, the number of features for each Task Contrast set of features.

HCP-A collected fMRI data from three tasks: Face Name (Sperling et al., 2001), Conditioned Approach Response Inhibition Task (CARIT) (Somerville et al., 2018) and VISual MOTOR (VISMOTOR) (Ances et al., 2009). First, the Face Name task (Sperling et al., 2001) taps into episodic memory. The task had three blocks. In the encoding block [Encoding], participants were asked to memorise the names of faces shown. These faces were then shown again in the recall block [Recall] when the participants were asked if they could remember the names of the previously shown faces. There was also the distractor block [Distractor]

occurring between the encoding and recall blocks. Here participants were distracted by a Go/NoGo task. We computed six contrasts for this Face Name task: [Encode], [Recall], [Distractor], [Encode vs. Distractor], [Recall vs. Distractor] and [Encode vs. Recall].

Second, the CARIT task (Somerville et al., 2018) was adapted from the classic Go/NoGo task and taps into inhibitory control. Participants were asked to press a button to all [Go] but not to two [NoGo] shapes. We computed three contrasts for the CARIT task: [NoGo], [Go] and [NoGo vs. Go].

Third, the VISMOTOR task (Ances et al., 2009) was designed to test simple activation of the motor and visual cortices. Participants saw a checkerboard with a red square either on the left or right. They needed to press a corresponding key to indicate the location of the red square. We computed just one contrast for the VISMOTOR task: [Vismotor], which indicates the presence of the checkerboard vs. baseline.

Sets of Features 11-13: Task fMRI functional connectivity (Task FC)

Task FC reflects functional connectivity (FC) among the brain regions during each task, which is considered an important source of individual differences (Elliott et al., 2019; Fair et al., 2007; Gratton et al., 2018). Unlike Task Contrasts, here we treated the double-gamma, convolved task events as regressors of no interest and focused on the residuals of the regression from each task (Fair et al., 2007). Using the same atlases as Task Contrast (Fischl et al., 2002; Glasser et al., 2016), we computed Pearson's correlations of each pair of 379 regions, resulting in a table of 71,631 non-overlapping FC indices for each task. We then applied r-to-z transformation and principal component analysis (PCA) of 75 components (Rasero et al., 2021; Sripada et al., 2019, 2020). Note to avoid data leakage, we conducted the PCA on each training set and applied its definition to the corresponding test set. Accordingly, there were three sets of 75 features for Task FC, one for each task.

Set of Features 14: Resting-state functional MRI functional connectivity (Rest FC)

Similar to Task FC, Rest FC reflects functional connectivity (FC) among the brain regions, except that Rest FC occurred during the resting (as opposed to task-performing) period. HCP-A collected Rest FC from four 6.42-min (488 frames) runs across two days, leading to 26-min long data (Harms et al., 2018). On each day, the study scanned two runs of Rest FC, starting with anterior-to-posterior (AP) and then with posterior-to-anterior (PA) phase encoding polarity. We used the "rfMRI_REST_Atlas_MSMAll_hp0_clean.dscalar.nii" file that was pre-processed and concatenated across the four runs. We applied the same computations (i.e., parcellation, Pearson's correlations, r-to-z transformation and PCA) with the Task FC.

Sets of Features 15-18: Structural MRI (sMRI)

sMRI reflects individual differences in brain anatomy. The HCP-A used an established pre-processing pipeline for sMRI (Glasser et al., 2013). We focused on four sets of features: cortical thickness, cortical surface area, subcortical volume and total brain volume. For cortical thickness and cortical surface area, we used Destrieux's atlas (Destrieux et al., 2010; Fischl, 2012) from FreeSurfer's "aparc.stats" file, resulting in 148 regions for each set of features. For subcortical volume, we used the aseg atlas (Fischl et al., 2002) from FreeSurfer's "aseg.stats" file, resulting in 19 regions. For total brain volume, we had five

FreeSurfer-based features: “FS_IntraCranial_Vol” or estimated intra-cranial volume, “FS_TotCort_GM_Vol” or total cortical grey matter volume, “FS_Tot_WM_Vol” or total cortical white matter volume, “FS_SubCort_GM_Vol” or total subcortical grey matter volume and “FS_BrainSegVol_eTIV_Ratio” or ratio of brain segmentation volume to estimated total intracranial volume.

Cognitive abilities: $\text{Cognition}_{\text{fluid}}$

We measured $\text{Cognition}_{\text{fluid}}$ via the NIH Toolbox (Weintraub et al., 2014), using the “fluidcogcomp_unadj” variable. $\text{Cognition}_{\text{fluid}}$ summarises scores from five tests assessed outside of the MRI: Dimensional Change Card Sort, Flanker Inhibitory Control and Attention, Picture Sequence Memory, List Sorting Working Memory and Pattern Comparison Processing Speed.

Prediction models for Brain Age and Brain Cognition

To compute Brain Age and Brain Cognition, we ran two separate prediction models. These prediction models either had chronological age or $\text{Cognition}_{\text{fluid}}$ as the target and standardised brain MRI as the features (Denissen et al., 2022). We used nested cross-validation (CV) to build these models. We first split the data into five folds (i.e., outer CV). One of these outer-CV folds was treated as a test set, and the rest was treated as a training set, which was further divided into five folds (i.e., inner CV). We used the inner CV to tune for hyperparameters of the models and the outer CV to evaluate the predictive performance of the models. We controlled for the potential influences of biological sex on the brain features by first residualising biological sex from brain features in the training set. We then applied the regression of this residualisation based on the training set to the test set. We also standardised the brain features in the training set and then used the mean and standard deviation of the training set to standardise the test set.

To demonstrate the predictive performance, we assessed the similarity between the observed values and the predicted values of each model across test sets, using Pearson’s *r*, coefficient of determination (R^2) and mean absolute error (MAE). Note that for R^2, we used the sum of squares definition (i.e., $R^2 = 1 – (\text{sum of squares residuals}/\text{total sum of squares})$) per a previous recommendation (Poldrack et al., 2020). We considered the predicted values from the test sets of models predicting age or $\text{Cognition}_{\text{fluid}}$, as Brain Age and Brain Cognition, respectively.

In addition to using each of the 18 sets of features in separate prediction models, we drew information across these sets via stacking. Specifically, we computed predicted values from each of the 18 sets of features in the training sets. We then treated different combinations of these predicted values as features to predict the targets in separate “stacked” models. We specified eight stacked models: “All” (i.e., including all 18 sets of features), “All excluding Task FC”, “All excluding Task Contrast”, “Non-Task” (i.e., including only Rest FC and sMRI), “Resting and Task FC”, “Task Contrast and FC”, “Task Contrast” and “Task FC”. Accordingly, in total, there were 26 prediction models for Brain Age and Brain Cognition.

For the machine learning algorithm, we used Elastic Net (Zou & Hastie, 2005) as implemented in sklearn (Pedregosa et al., 2011). Not only does Elastic Net provide a relatively good predictive performance for fMRI (Dubois et al., 2018; Pat, Wang, Anney, et

al., 2022; Pat, Wang, Bartonicek, et al., 2022; Tetereva et al., 2022), but it also offers easy-to-interpret feature importance (Molnar, 2019). In our grid search, we tuned two Elastic Net hyperparameters: α using 70 numbers in log space, ranging from .1 and 100, and ℓ_1-ratio using 25 numbers in linear space, ranging from 0 and 1.

Brain Age calculations: Brain Age, Brain Age Gap, Corrected Brain Age and Corrected Brain Age Gap

In addition to Brain Age, which is the predicted value from the models predicting chronological age in the test sets, we calculated three other indices to reflect the estimation of brain aging. First, Brain Age Gap reflects the difference between the age predicted by brain MRI and the actual, chronological age. Here we simply subtracted the chronological age from Brain Age.

Next, to correct for the estimation biases (i.e., overestimation in younger, but underestimation in older, individuals (Le et al., 2018)), we applied de Lange and Cole's (2020) 5th equation. Specifically, we first fit a regression line predicting the Brain Age from a chronological age in a training set. We then used the slope and intercept of this regression line to adjust Brain Age in a test set, resulting in Corrected Brain Age.

Lastly, we computed Corrected Brain Age Gap by subtracting the chronological age from the Corrected Brain Age (Butler et al., 2021; Le et al., 2018).

The ability of Brain Age indices to capture Cognition$_{fluid}$

We first combined Brain Age, Brain Cognition, chronological age and Cognition$_{fluid}$ across test sets into one table. We then conducted three sets of regression analyses to demonstrate the ability of different Brain Age indices, calculated from 26 different prediction models based on different sets of brain MRI features, to capture Cognition$_{fluid}$.

1. <u>Simple Regression: Using each Brain Age index to explain Cognition$_{fluid}$</u>

Here using simple regression, we simply had each Brain Age index as the sole regressor for Cognition$_{fluid}$:

$$\text{Cognition}_{fluid} \sim \beta_0 + \beta_1 \text{ Brain Age Index}_i \text{ , (1)}$$

where i is the index for the four Brain Age indices. Because different Brain Age indices differ in the adjustments applied, this simple regression could reveal the extent to which each adjustment influences variation in Cognition$_{fluid}$ explained. Additionally, Brain Age calculated from 26 different prediction models would have different levels of predictive performance in predicting chronological age. Accordingly, this simple regression could also reveal if Brain Age from a better-performing age-prediction model was able to capture more variation in Cognition$_{fluid}$.

In addition to Brain Age indices, we also used simple regression to test how well Brain Cognition as a sole regressor explains Cognition$_{fluid}$:

$Cognition_{fluid} \sim \beta_0 + \beta_1$ Brain Cognition, (2)

This allows us to compare the ability of Brain Age Indices vs. Brain Cognition as a sole regressor for predicting $Cognition_{fluid}$.

2. Multiple Regression: Using chronological age and each Brain Age index to explain $Cognition_{fluid}$

Here using multiple regression, we had both chronological age and each Brain Age index as the regressors for $Cognition_{fluid}$:

$Cognition_{fluid} \sim \beta_0 + \beta_1$ Chonological Age + β_2 Brain Age $Index_i$, (3)

Having chronological age in the same regression model as a Brain Age index allowed us to control for the effects of chronological age on the Brain Age index, thereby, revealing the unique effects of the Brain Age index (Butler et al., 2021; Le et al., 2018). To formally determine the unique effects of a Brain Age index on $Cognition_{fluid}$ along with the effects it shared with chronological age (i.e., common effects), we applied the commonality analysis (Nimon et al., 2008). For the unique effects, we computed ΔR^2. ΔR^2 is the increase in R^2 when having an additional regressor in the regression model:

Unique $Effects_{chronological\ age} = \Delta R^2_{chronological\ age} = R^2_{chronological\ age,\ Brain\ Age\ index} - R^2_{Brain\ Age\ index}$
Unique $Effects_{Brain\ Age\ index} = \Delta R^2_{Brain\ Age\ index} = R^2_{chronological\ age,\ Brain\ Age\ index} - R^2_{chronological\ age}$, (4)

We determined the statistical significance of ΔR^2 by:

$$F_{Change} = \frac{(N-k_2-1)\Delta R^2}{k_{change}(1-R_2^2)}, (5)$$

where F_{Change} is the F-ratio (with the degree of freedom of k_{Change} and $N - k_2 - 1$), N is the number of observations, $_2$ is the model with more regressors, k is the number of regressors, k_{Change} is the difference between the number of regressors.

As for the common effects between chronological age and each Brain Age index, we used the below calculation:

Common $Effects_{chronological\ age,\ Brain\ Age\ index} = R^2_{chronological\ age,\ Brain\ Age\ index} - \Delta R^2_{chronological\ age} - \Delta R^2_{Brain\ Age\ index}$, (6)

These common effects indicate the extent to which variation in $Cognition_{fluid}$ explained by each Brain Age index was shared with chronological age. Note the common effects can be negative, especially with a high multicollinearity (Ray-Mukherjee et al., 2014). To deal with this, we treated negative common effects as zero (Frederick, 1999) and then scaled variation explained by other effects to be proportional to the total effects of the full regression model.

3. <u>Multiple Regression: Using chronological age, each Brain Age index and Brain Cognition to explain $\text{Cognition}_{\text{fluid}}$</u>

Similar to the above multiple regression model, we had chronological age, each Brain Age index and Brain Cognition as the regressors for $\text{Cognition}_{\text{fluid}}$:

$$\text{Cognition}_{\text{fluid}} \sim \beta_0 + \beta_1 \text{Chronological Age} + \beta_2 \text{Brain Age Index}_i + \beta_3 \text{Brain Cognition}, \quad (7)$$

Applying the commonality analysis here allowed us, first, to investigate the addictive, unique effects of Brain Cognition, over and above chronological age and Brain Age indices. More importantly, the commonality analysis also enabled us to test the common, shared effects that Brain Cognition had with chronological age and Brain Age indices in explaining $\text{Cognition}_{\text{fluid}}$. We calculated the commonality analysis as follows (Nimon et al., 2017):

$$\text{Unique Effects}_{\text{chronological age}} = \Delta R^2_{\text{chronological age}} = R^2_{\text{chronological age, Brain Age index, Brain Cognition}} - R^2_{\text{Brain Age index, Brain Cognition}}$$

$$\text{Unique Effects}_{\text{Brain Age index}} = \Delta R^2_{\text{Brain Age index}} = R^2_{\text{chronological age, Brain Age index, Brain Cognition}} - R^2_{\text{chronological age, Brain Cognition}}$$

$$\text{Unique Effects}_{\text{Brain Cognition}} = \Delta R^2_{\text{Brain Cognition}} = R^2_{\text{chronological age, Brain Age index, Brain Cognition}} - R^2_{\text{chronological age, Brain Age Index}}$$

$$\text{Common Effects}_{\text{chronological age, Brain Age index}} = R^2_{\text{chronological age, Brain Cognition}} + R^2_{\text{Brain Age index, Brain Cognition}} - R^2_{\text{Brain Cognition}} - R^2_{\text{chronological age, Brain Age index, Brain Cognition}}$$

$$\text{Common Effects}_{\text{chronological age, Brain Cognition}} = R^2_{\text{chronological age, Brain Age Index}} + R^2_{\text{Brain Age index, Brain Cognition}} - R^2_{\text{Brain Age Index}} - R^2_{\text{chronological age, Brain Age index, Brain Cognition}}$$

$$\text{Common Effects}_{\text{Brain Age index, Brain Cognition}} = R^2_{\text{chronological age, Brain Age Index}} + R^2_{\text{chronological age, Brain Cognition}} - R^2_{\text{chronological age}} - R^2_{\text{chronological age, Brain Age index, Brain Cognition}}$$

$$\text{Common Effects}_{\text{chronological age, Brain Age index, Brain Cognition}} = R^2_{\text{chronological age}} + R^2_{\text{Brain Age Index}} + R^2_{\text{Brain Cognition}} - R^2_{\text{chronological age, Brain Age Index}} - R^2_{\text{chronological age, Brain Cognition}} - R^2_{\text{Brain Age Index, Brain Cognition}} - R^2_{\text{chronological age, Brain Age index, Brain Cognition}}, \quad (8)$$

Code Accessibility

The shell and Python scripts used in the analyses are made available here: https://github.com/HAM-lab-Otago-University/HCP-Aging_commonality

References

Ances, B. M., Liang, C. L., Leontiev, O., Perthen, J. E., Fleisher, A. S., Lansing, A. E., & Buxton, R. B. (2009). Effects of aging on cerebral blood flow, oxygen metabolism, and blood oxygenation level dependent responses to visual stimulation. *Human Brain Mapping*, *30*(4), 1120–1132. https://doi.org/10.1002/hbm.20574

Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C., & Mechelli, A. (2021). Machine learning for brain age prediction: Introduction to methods and clinical applications. *EBioMedicine*, *72*, 103600. https://doi.org/10.1016/j.ebiom.2021.103600

Beheshti, I., Nugent, S., Potvin, O., & Duchesne, S. (2019). Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. *NeuroImage: Clinical*, *24*, 102063. https://doi.org/10.1016/j.nicl.2019.102063

Bookheimer, S. Y., Salat, D. H., Terpstra, M., Ances, B. M., Barch, D. M., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Diaz-Santos, M., Elam, J. S., Fischl, B., Greve, D. N., Hagy, H. A., Harms, M. P., Hatch, O. M., Hedden, T., Hodge, C., Japardi, K. C., Kuhn, T. P., … Yacoub, E. (2019). The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage*, *185*, 335–348. https://doi.org/10.1016/j.neuroimage.2018.10.009

Boyle, R., Jollans, L., Rueda-Delgado, L. M., Rizzo, R., Yener, G. G., McMorrow, J. P., Knight, S. P., Carey, D., Robertson, I. H., Emek-Savaş, D. D., Stern, Y., Kenny, R. A., & Whelan, R. (2021). Brain-predicted age difference score is related to specific cognitive functions: A multi-site replication analysis. *Brain Imaging and Behavior*, *15*(1), 327–345. https://doi.org/10.1007/s11682-020-00260-3

Butler, E. R., Chen, A., Ramadan, R., Le, T. T., Ruparel, K., Moore, T. M., Satterthwaite, T. D., Zhang, F., Shou, H., Gur, R. C., Nichols, T. E., & Shinohara, R. T. (2021). Pitfalls in brain age analyses. *Human Brain Mapping*, *42*(13), 4092–4101. https://doi.org/10.1002/hbm.25533

Cole, J. H. (2020). Multimodality neuroimaging brain-age in UK biobank: Relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of Aging*, *92*, 34–42. https://doi.org/10.1016/j.neurobiolaging.2020.03.014

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S. E., Zhang, Q., Wray, N. R., Redmond, P., Marioni, R. E., Starr, J. M., Cox, S. R., Wardlaw, J. M., Sharp, D. J., & Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, *23*(5), Article 5. https://doi.org/10.1038/mp.2017.62

de Lange, A.-M. G., & Cole, J. H. (2020). Commentary: Correction procedures in brain-age prediction. *NeuroImage: Clinical*, *26*, 102229. https://doi.org/10.1016/j.nicl.2020.102229

Denissen, S., Engemann, D. A., De Cock, A., Costers, L., Baijot, J., Laton, J., Penner, I., Grothe, M., Kirsch, M., D'hooghe, M. B., D'Haeseleer, M., Dive, D., De Mey, J., Van Schependom, J., Sima, D. M., & Nagels, G. (2022). Brain age as a surrogate marker for cognitive performance in multiple sclerosis. *European Journal of Neurology*, *29*(10), 3039–3049. https://doi.org/10.1111/ene.15473

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756), 20170284. https://doi.org/10.1098/rstb.2017.0284

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., Ireland, D., Ramrakha, S., Poulton, R., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, *189*, 516–532. https://doi.org/10.1016/j.neuroimage.2019.01.068

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaître, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *ELife*, *9*, e54055. https://doi.org/10.7554/eLife.54055

Fair, D. A., Schlaggar, B. L., Cohen, A. L., Miezin, F. M., Dosenbach, N. U. F., Wenger, K. K., Fox, M. D., Snyder, A. Z., Raichle, M. E., & Petersen, S. E. (2007). A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. *NeuroImage*, *35*(1), 396–405. https://doi.org/10.1016/j.neuroimage.2006.11.051

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole Brain Segmentation. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Franke, K., & Gaser, C. (2019). Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Frontiers in Neurology*, *10*, 789. https://doi.org/10.3389/fneur.2019.00789

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, *50*(3), 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. *Advances in Social Science Methodology*, *5*, 305–318.

Frisoni, G. B., Boccardi, M., Barkhof, F., Blennow, K., Cappa, S., Chiotis, K., Démonet, J.-F., Garibotto, V., Giannakopoulos, P., Gietl, A., Hansson, O., Herholz, K., Jack, C. R., Nobili, F., Nordberg, A., Snyder, H. M., Kate, M. T., Varrone, A., Albanese, E., … Winblad, B. (2017). Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *The Lancet Neurology*, *16*(8), 661–676. https://doi.org/10.1016/S1474-4422(17)30159-X

Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., Coalson, T. S., Harms, M. P., Jenkinson, M., Moeller, S., Robinson, E. C., Sotiropoulos, S. N., Xu, J., Yacoub, E., Ugurbil, K., & Van Essen, D. C. (2016). The Human Connectome Project's neuroimaging approach. *Nature Neuroscience*, *19*(9), 1175–1187. https://doi.org/10.1038/nn.4361

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013).

The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, *26*(1), 288–303. https://doi.org/10.1093/cercor/bhu239

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, *98*(2), 439-452.e5. https://doi.org/10.1016/j.neuron.2018.03.035

Harms, M. P., Somerville, L. H., Ances, B. M., Andersson, J., Barch, D. M., Bastiani, M., Bookheimer, S. Y., Brown, T. B., Buckner, R. L., Burgess, G. C., Coalson, T. S., Chappell, M. A., Dapretto, M., Douaud, G., Fischl, B., Glasser, M. F., Greve, D. N., Hodge, C., Jamison, K. W., … Yacoub, E. (2018). Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage*, *183*, 972–984. https://doi.org/10.1016/j.neuroimage.2018.09.060

Jiang, R., Scheinost, D., Zuo, N., Wu, J., Qi, S., Liang, Q., Zhi, D., Luo, N., Chung, Y., Liu, S., Xu, Y., Sui, J., & Calhoun, V. (2022). A Neuroimaging Signature of Cognitive Aging from Whole-Brain Functional Connectivity. *Advanced Science*, *9*(24), 2201621. https://doi.org/10.1002/advs.202201621

Jonsson, B. A., Bjornsdottir, G., Thorgeirsson, T. E., Ellingsen, L. M., Walters, G. B., Gudbjartsson, D. F., Stefansson, H., Stefansson, K., & Ulfarsson, M. O. (2019). Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-13163-9

Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., Paulus, M. P., Tulsa 1000 Investigators, Aupperle, R. L., Bodurka, J., Cha, Y.-H., Feinstein, J. S., Khalsa, S. S., Savitz, J., Simmons, W. K., & Victor, T. A. (2018). A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in Aging Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnagi.2018.00317

Liang, H., Zhang, F., & Niu, X. (2019). Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Human Brain Mapping*, *40*(11), 3143–3152. https://doi.org/10.1002/hbm.24588

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J. M., Lampe, L., Rahim, M., Abraham, A., Craddock, R. C., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M. L., Witte, A. V., Villringer, A., & Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, *148*, 179–188. https://doi.org/10.1016/j.neuroimage.2016.11.005

Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466. https://doi.org/10.3758/BRM.40.2.457

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2017). Erratum to: An R package to compute commonality coefficients in the multiple regression case: An introduction

to the package and a practical example. *Behavior Research Methods*, *49*(6), 2275–2275. https://doi.org/10.3758/s13428-017-0853-2

Pat, N., Wang, Y., Anney, R., Riglin, L., Thapar, A., & Stringaris, A. (2022). Longitudinally stable, brain-based predictive models mediate the relationships between childhood cognition and socio-demographic, psychological and genetic factors. *Human Brain Mapping*, hbm.26027. https://doi.org/10.1002/hbm.26027

Pat, N., Wang, Y., Bartonicek, A., Candia, J., & Stringaris, A. (2022). Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cerebral Cortex*, bhac235. https://doi.org/10.1093/cercor/bhac235

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, *77*(5), 534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671

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

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328. https://doi.org/10.1111/2041-210X.12166

Richard, G., Kolskår, K., Sanders, A.-M., Kaufmann, T., Petersen, A., Doan, N. T., Sánchez, J. M., Alnæs, D., Ulrichsen, K. M., Dørum, E. S., Andreassen, O. A., Nordvik, J. E., & Westlye, L. T. (2018). Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ*, *6*, e5908. https://doi.org/10.7717/peerj.5908

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P. V., Van Essen, D. C., Smith, S. M., Edwards, A. D., … Rueckert, D. (2018). Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*, *167*, 453–465. https://doi.org/10.1016/j.neuroimage.2017.10.037

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, *34*(1), 17–24. https://doi.org/10.1037/pag0000288

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of brain age delta from brain imaging. *NeuroImage*, *200*, 528–539. https://doi.org/10.1016/j.neuroimage.2019.06.017

Somerville, L. H., Bookheimer, S. Y., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Dapretto, M., Elam, J. S., Gaffrey, M. S., Harms, M. P., Hodge, C., Kandala, S., Kastman, E. K., Nichols, T. E., Schlaggar, B. L., Smith, S. M., Thomas, K. M., Yacoub, E., Van Essen, D. C., & Barch, D. M. (2018). The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage*, *183*, 456–468. https://doi.org/10.1016/j.neuroimage.2018.08.050

Sperling, R. A., Bates, J. F., Cocchiarella, A. J., Schacter, D. L., Rosen, B. R., & Albert, M. S. (2001). Encoding novel face-name associations: A functional MRI study. *Human Brain Mapping*, *14*(3), 129–139. https://doi.org/10.1002/hbm.1047

Sripada, C., Angstadt, M., Rutherford, S., Kessler, D., Kim, Y., Yee, M., & Levina, E. (2019). Basic Units of Inter-Individual Variation in Resting State Connectomes. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-018-38406-5

Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., & Shedden, K. (2020). Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping*, *41*(12), 3186–3197. https://doi.org/10.1002/hbm.25007

Tetereva, A., Li, J., Deng, J. D., Stringaris, A., & Pat, N. (2022). Capturing brain-cognition relationship: Integrating task-based fMRI across tasks markedly boosts prediction and test-retest reliability. *NeuroImage*, *263*, 119588. https://doi.org/10.1016/j.neuroimage.2022.119588

Vieira, B. H., Pamplona, G. S. P., Fachinello, K., Silva, A. K., Foss, M. P., & Salmon, C. E. G. (2022). On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence*, *93*, 101654. https://doi.org/10.1016/j.intell.2022.101654

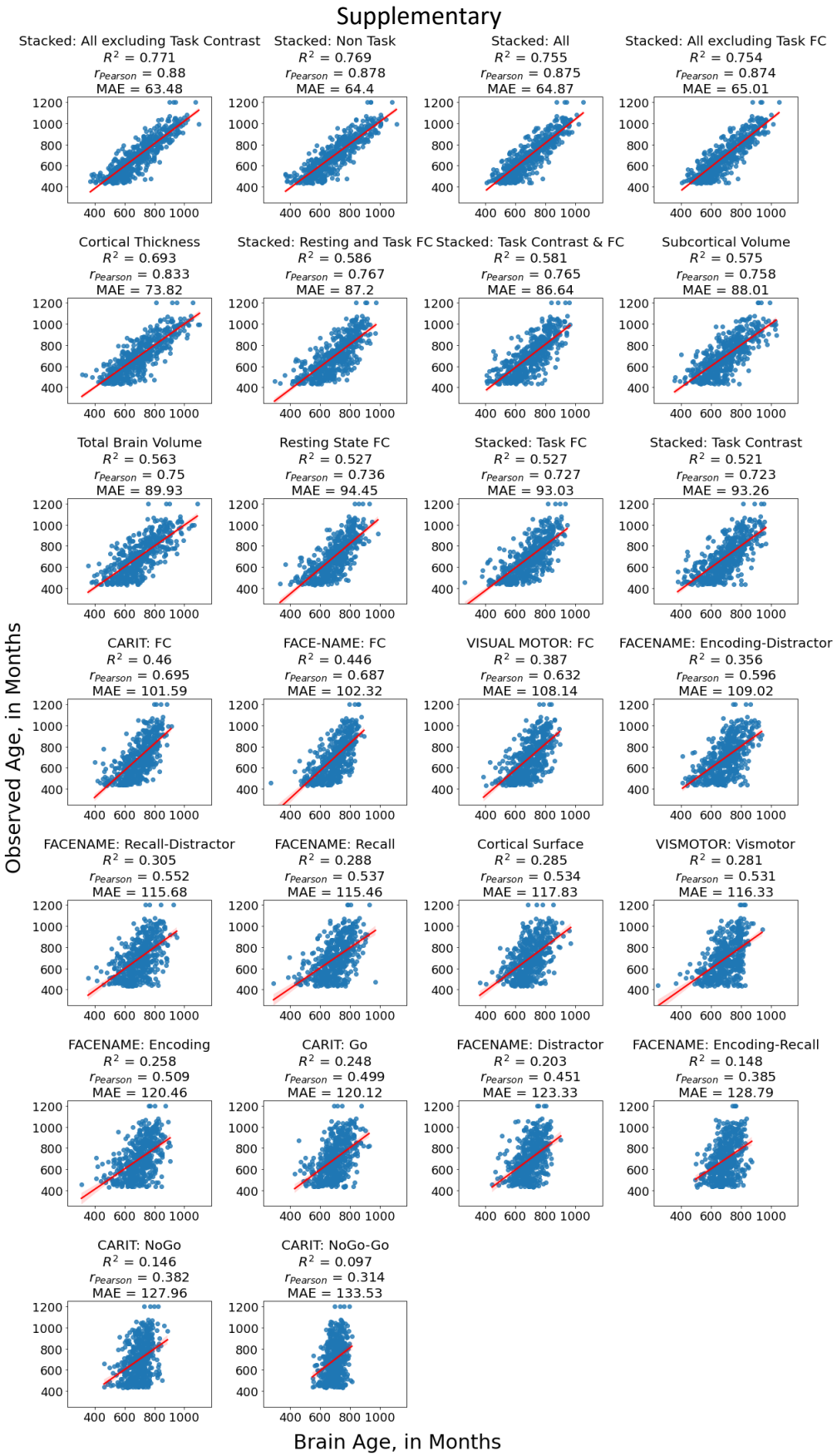
Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Slotkin, J., Carlozzi, N. E., Bauer, P. J., Wallner-Allen, K., Fox, N., Havlik, R., Beaumont, J. L., Mungas, D., Manly, J. J., Moy, C., Conway, K., Edwards, E., Nowinski, C. J., & Gershon, R. (2014). The Cognition Battery of the NIH Toolbox for Assessment of Neurological and Behavioral Function: Validation in an Adult Sample. *Journal of the International Neuropsychological Society*, *20*(6), 567–578. https://doi.org/10.1017/S1355617714000320

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

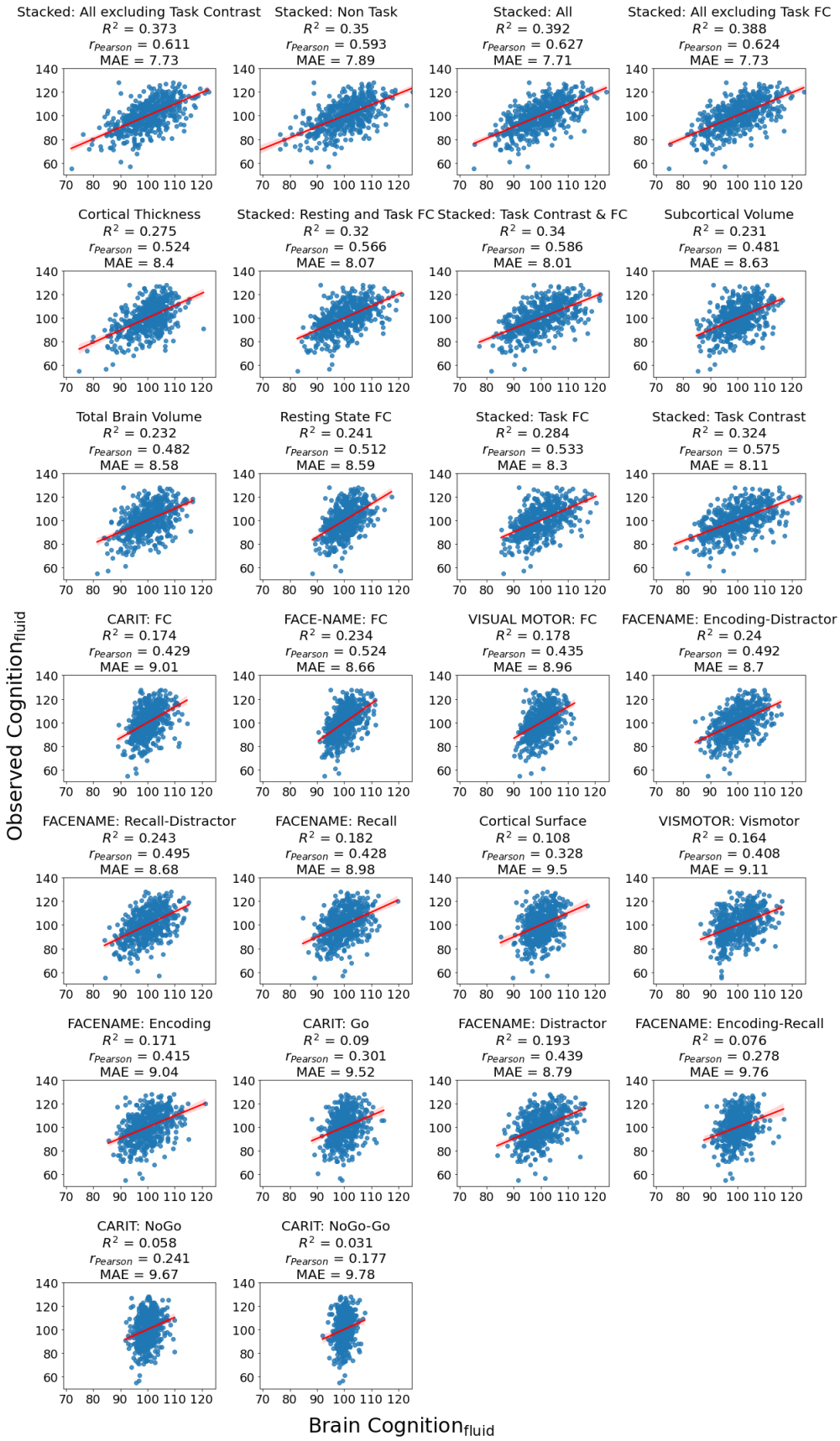
Wrigglesworth, J., Yaacob, N., Ward, P., Woods, R. L., McNeil, J., Storey, E., Egan, G., Murray, A., Shah, R. C., Jamadar, S. D., Trevaks, R., Ward, S., Harding, I. H., & Ryan, J. (2022). Brain-predicted age difference is associated with cognitive processing in later-life. *Neurobiology of Aging*, *109*, 195–203. https://doi.org/10.1016/j.neurobiolaging.2021.10.007

Zhao, X., & Zhao, X.-M. (2021). Deep learning of brain magnetic resonance images: A brief review. *Methods*, *192*, 131–140. https://doi.org/10.1016/j.ymeth.2020.09.007

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x



Supplementary Figure 1. The scatter plots between observed and predicted values in the test sets from age-prediction models.



Supplementary Figure 2. The scatter plots between observed and predicted values in the test sets from cognition-prediction models