Brain age has limited utility as a biomarker for capturing fluid cognition in older individuals

Alina Tetereva, Narun Pat*

Department of Psychology, University of Otago, Dunedin, New Zealand

Abstract One well-known biomarker candidate that supposedly helps capture fluid cognition is Brain Age, or a predicted value based on machine-learning models built to predict chronological age from brain MRI. To formally evaluate the utility of Brain Age for capturing fluid cognition, we built 26 age-prediction models for Brain Age based on different combinations of MRI modalities, using the Human Connectome Project in Aging (n=504, 36–100 years old). First, based on commonality analyses, we found a large overlap between Brain Age and chronological age: Brain Age could uniquely add only around 1.6% in explaining variation in fluid cognition over and above chronological age. Second, the age-prediction models that performed better at predicting chronological age did NOT necessarily create better Brain Age for capturing fluid cognition 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 fluid cognition. Third, Brain Age missed around 11% of the total variation in fluid cognition that could have been explained by the brain variation. That is, directly predicting fluid cognition from brain MRI data (instead of relying on Brain Age and chronological age) could lead to around a 1/3-time improvement of the total variation explained. Accordingly, we demonstrated the limited utility of Brain Age as a biomarker for fluid cognition and made some suggestions to ensure the utility of Brain Age in explaining fluid cognition and other phenotypes of interest.

eLife assessment

This **useful** manuscript challenges the utility of current paradigms for estimating brain-age with magnetic resonance imaging measures. It presents **solid** evidence to support the suggestion that an alternative approach focused on predicting cognition may be more beneficial. This work will be of interest to researchers working on brain-age and related models.

Introduction

Older adults often experience declines in several cognitive abilities such as memory, attention and processing speed, collectively known as fluid cognition (*Salthouse, 2019*; *Weintraub et al., 2014*). Having objective biomarkers to capture fluid cognition 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 fluid cognition in older adults (*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 fluid cognition, we still need to address at least the three impeding issues.

First, to what extent does having information on Brain Age improve its utility to capture fluid cognition in older adults beyond knowing a person's chronological age? To compute Brain Age, researchers first build a prediction model that predicts the chronological age based on a person's brain MRI 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 fluid cognition, 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 and 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). Moreover, Brain Age is known to demonstrate the 'regression towards the mean' phenomenon (*Stigler, 1997*). More specifically, because Brain Age is a predicted value of a regression model that predicts chronological age, Brain Age is usually shrunk towards the mean age of samples used for training the model (*Butler et al., 2021*; *de Lange and Cole, 2020*; *Le et al., 2018*). Accordingly, Brain Age predicts chronological age more accurately for individuals who are closer to the mean age while overestimating younger individuals' chronological age and underestimating older individuals' chronological age. There are many adjustments proposed to correct for the age dependency, but the outcomes tend to be similar to each other (*Beheshti et al., 2019*; *de Lange and 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 age dependency (*Butler et al., 2021*). Here, we tested the utility of different Brain Age calculations in capturing fluid cognition, over and above chronological age.

Second, do better-performing age-prediction models correspond to the improvement in the utility to capture fluid cognition? 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 and Zhao, 2021*). However, the improvement in predicting chronological age may not necessarily make Brain Age Gap better at capturing fluid cognition (*Jirsaraie et al., 2023a*). 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 fluid cognition beyond chronological age. Here we examined the performance of age-prediction models as a function of their utility in capturing fluid cognition.

Third and finally, certain variation in fluid cognition is related to brain MRI, but to what extent does Brain Age not capture this variation? To estimate the variation in fluid cognition that is related to brain MRI, we could build prediction models that directly predict fluid cognition (i.e. as opposed to chronological age) from brain MRI data. Previous studies found reasonable predictive performances of these cognition-prediction models, built from certain MRI modalities (*Dubois et al., 2018*; *Pat et al., 2022*; *Rasero et al., 2021*; *Sripada et al., 2020*; *Tetereva et al., 2022*; for review, see *Vieira et al., 2022*). Analogous to Brain Age, we called the predicted values from these cognition-prediction models, Brain Cognition. The strength of an out-of-sample relationship between Brain Cognition and fluid cognition reflects variation in fluid cognition that is related to the brain MRI and, therefore, indicates the upper limit of Brain Age's capability in capturing fluid cognition. This is, by design, the variation in fluid cognition explained by Brain Cognition should be higher or equal to that explained by Brain Age. Consequently, if we included Brain Cognition, Brain Age and chronological age in the same model to explain fluid cognition, we would be able to examine the unique effects of Brain Cognition that explain fluid cognition beyond Brain Age and chronological age. These unique effects of Brain Cognition, in turn, would indicate the amount of co-variation between brain MRI and fluid cognition that is missed by Brain Age.

Our study set out to test the utility of Brain Age as a biomarker for capturing variation in fluid cognition among aging individuals. Using aging participants (36–100 years old) from the Human Connectome Project in Aging (*Bookheimer et al., 2019*), 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' utility in explaining fluid cognition in unseen participants. To test this utility of Brain Age indices, we applied simple regression models with each Brain Age index as a sole regressor to explain fluid cognition. Next, to test the unique effects of Brain Age in explaining fluid cognition beyond chronological age, we applied multiple regression models with both each Brain Age index and chronological age as regressors to explain fluid cognition. To reveal how much chronological age and Brain Age indices had in common in explaining fluid cognition (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 utility of Brain Age indices in explaining fluid cognition over and above lower-performing age-prediction models. Finally, we investigated the extent to which Brain Age indices missed the variation in fluid cognition that could be explained by the brain MRI. Here, we tested Brain Cognition's unique effects in multiple regression models with a Brain Age index, chronological age and Brain Cognition as regressors to explain fluid cognition.

Results

Relationship between chronological age and fluid cognition

Figure 1a shows the negative relationship between chronological age and fluid cognition (r(502) = –0.57, p<0.001, R^2=0.32). Older individuals tended to have a lower fluid cognition score.

Predictive performance of prediction models for Brain Age and Brain Cognition

Figure 1b and c show the predictive performance of different sets of brain MRI features in predicting chronological age and fluid cognition, respectively. For age prediction, the top-four models that performed similarly 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 >0.76, r>0.83, MAE <65 months). For fluid cognition prediction, the top-performing model was 'Stacked: All' (R^2=0.393, r=0.627, MAE = 7.7 points). The best set of features across age and fluid cognition prediction was cortical thickness. Across sets of MRI features, the age-prediction models tended to provide higher R^2 and r than the fluid cognition-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 4* shows the stability of feature importance across different outer-fold test sets.

Simple regression: using each Brain Age index to explain fluid cognition

Figure 5a shows variation in fluid cognition 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 fluid cognition. However, Brain Age Gap created from the *lower*-performing age-prediction models explained a higher amount of variation in fluid cognition. 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 fluid cognition, but, at the same time, produced Brain Age Gap that explained the least amount of variation in fluid cognition.

On the contrary, an amount of variation in fluid cognition explained by Corrected Brain Age Gap was relatively small (maximum at R^2=0.041) across age-prediction models and did not relate to the predictive performance of the age-prediction models. *Figure 5b* shows variation in fluid cognition explained by Brain Cognition, as compared to Brain Age indices. Brain Cognition appeared to explain
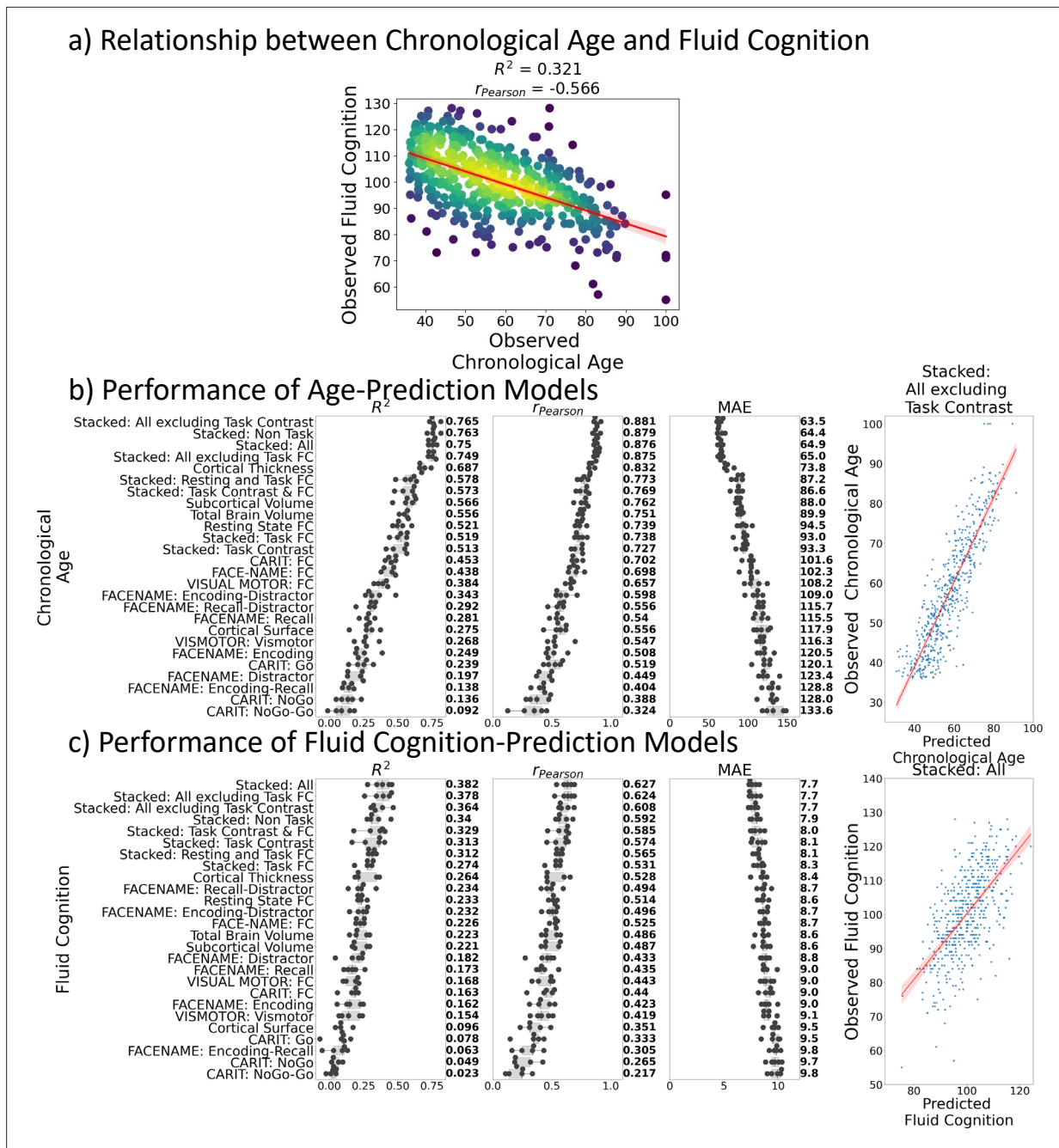
Figure 1. Relationship between chronological age and fluid cognition. (**a**) and predictive performance of prediction models using Brain MRI from different sets of MRI features to predict chronological age (**b**) and fluid cognition (**c**). Each dot in (**b**) and (**c**) represents predictive performance at each of the five outer-fold test sets. The numbers to the right of the predictive performance plots indicate the mean of predictive performance across the five outer-fold test sets. Note we only provided the scatter plots between observed and predicted values in the outer-fold test sets from the best prediction models for each target (age in years and fluid cognition in points) in this figure. See *Figure 2—figure supplement 1* and *Figure 2—figure supplement 2* for the scatter plots from other prediction models.
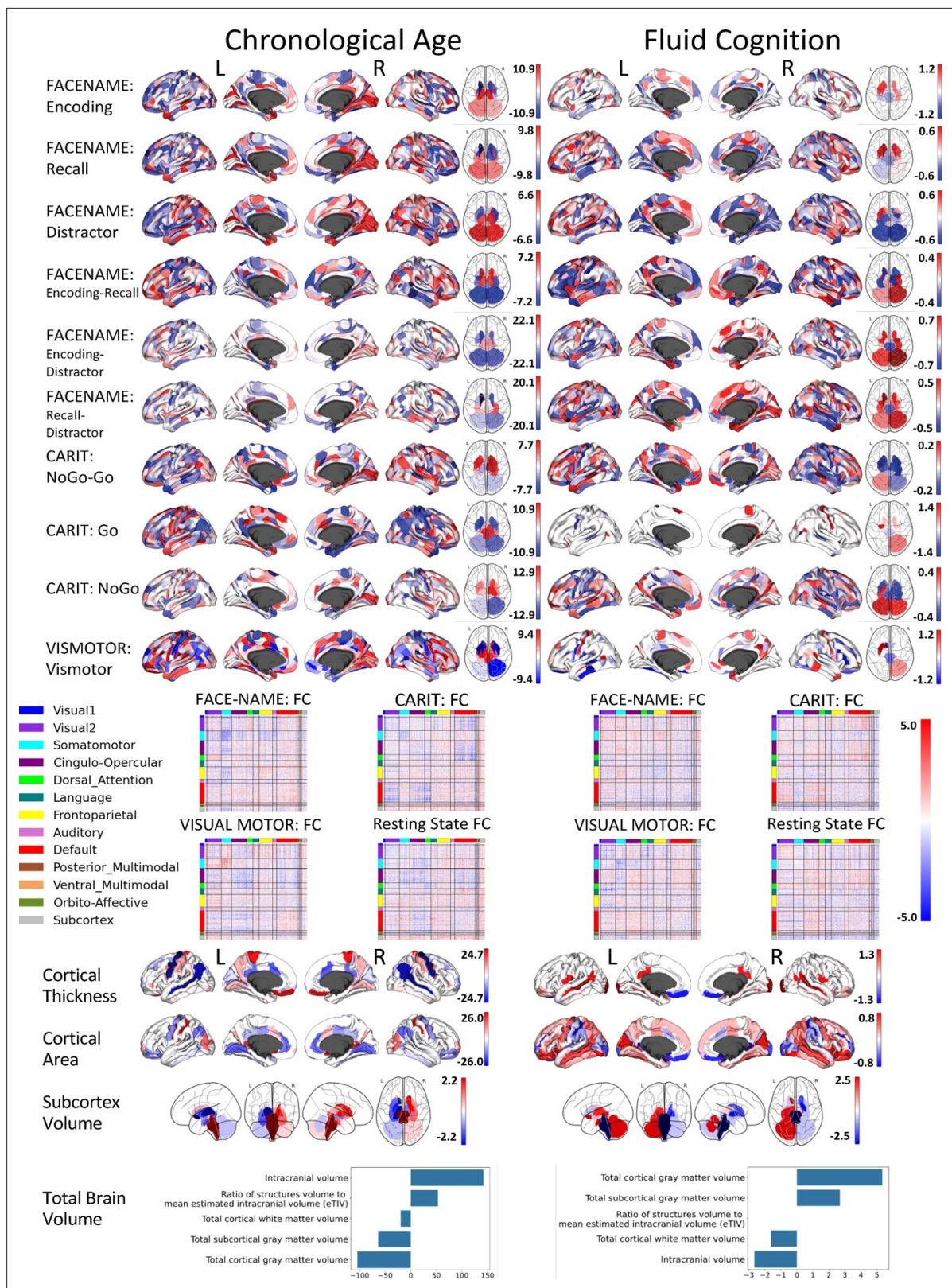
Figure 2. Feature importance (i.e. Elastic Net Coefficients) of prediction models based on each of the 18 sets of features.

The online version of this article includes the following figure supplement(s) for figure 2:

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

Figure supplement 2. The scatter plots between observed and predicted values in the outer-fold test sets from cognition-prediction models.
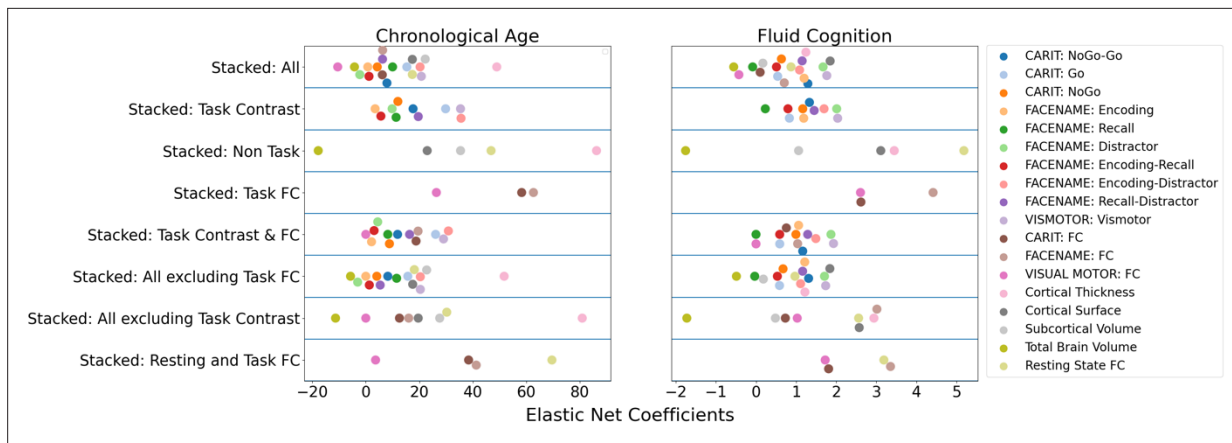
Figure 3. Feature importance (i.e. Elastic Net Coefficients) of the eight stacked prediction models.

a higher amount of variation in fluid cognition than any Brain Age indices, especially for top-performing age/cognition-prediction models (e.g. Stacked: All).

Multiple regression: Using chronological age and each Brain Age index to explain fluid cognition

Figure 6 shows the commonality analysis of multiple regression models, having both chronological age and each Brain Age index as the regressors for fluid cognition. We found R^2 for these models at M=0.326 (SD = 0.005). The unique effects of Brain Age indices were all relatively small (maximum at $\Delta R^2_{Brain\ Age\ index}$ = 0.0161, with statistically significant at p-value <0.05 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
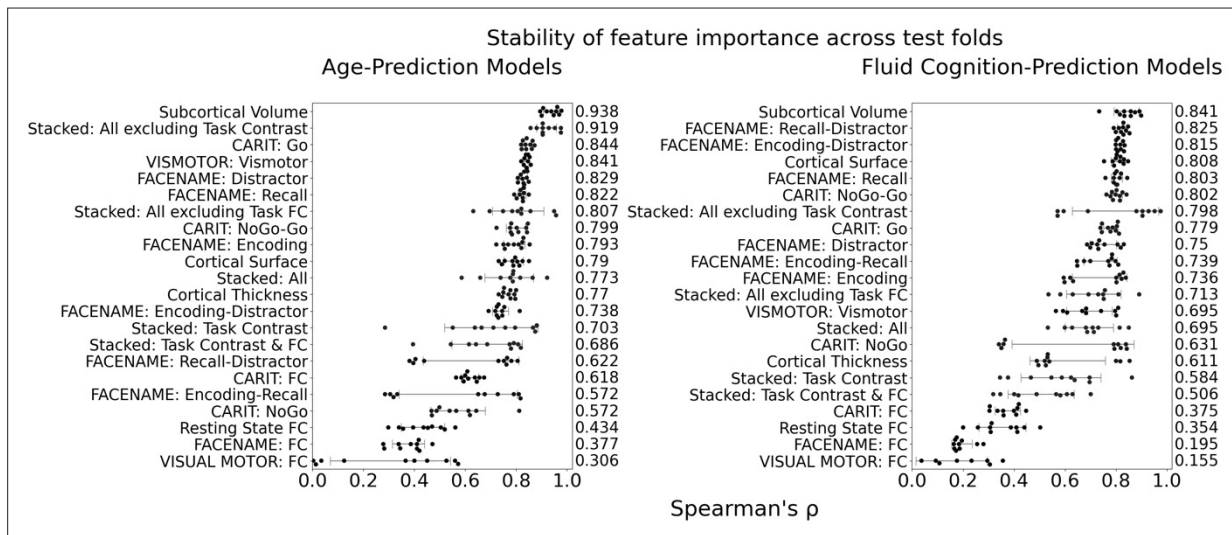

Figure 4. Stability of feature importance (i.e. Elastic Net Coefficients) of prediction models. Each dot represents rank stability (reflected by Spearman's ρ) in the feature importance between two prediction models of the same features, used in two different outer-fold test sets. Given that there were five outer-fold test sets, there were 10 Spearman's ρs for each prediction model. The numbers to the right of the plots indicate the mean of Spearman's ρ for each prediction model.
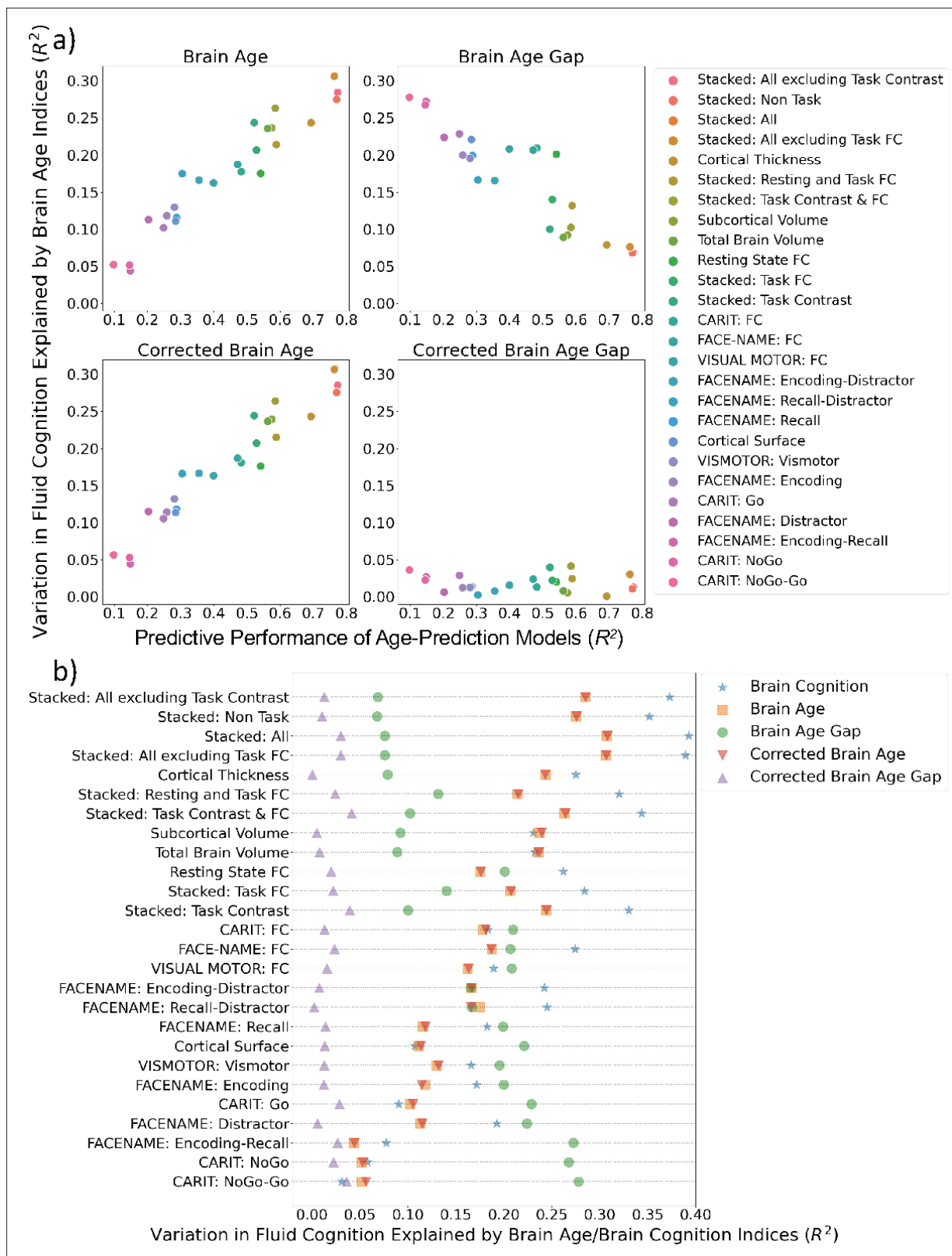
Figure 5. Simple regression: using each Brain Age index or Brain Cognition to explain fluid cognition. (**a**) shows variation in fluid cognition explained by each Brain Age index as a function of the predictive performance of age-prediction models. (**b**) plots variation in fluid cognition explained by Brain Age indices and Brain Cognition.
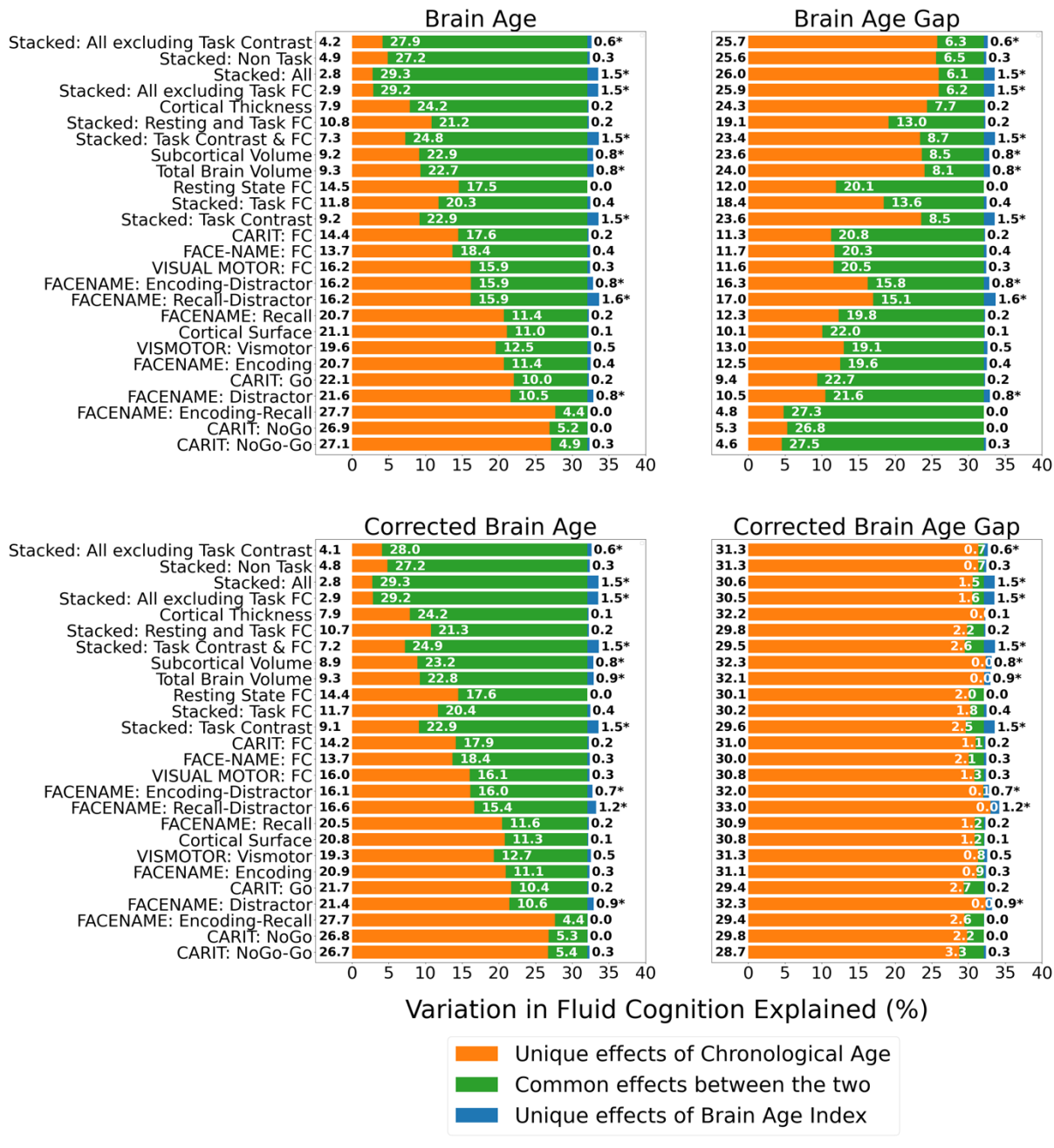
Figure 6. Commonality analysis of multiple regressions, having both chronological age and each Brain Age index as the regressors for capturing fluid cognition. The numbers to the left of the figure represent the unique effects of chronological age in %, the numbers in the middle of the figure represent the common effects between chronological age and Brain Age index in %, and the numbers to the right of the figure represent the unique effects of Brain Age Index in %. * represents the statistical significance of the unique effects of Brain Age Index at p<0.05.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Commonality analysis of Ridge regressions, having both chronological age and each Brain Age index as the regressors for capturing fluid cognition.

Figure supplement 2. Commonality analysis of Ridge regressions, having chronological age and each Brain Age index and Brain Cognition as the regressors for fluid cognition.

Age Gap, the unique effects of chronological age were much higher than the common effects across all age-prediction models.

Multiple regression: Using chronological age, each Brain Age index and Brain Cognition to explain fluid cognition

Figure 7 shows the commonality analysis of multiple regression models, having chronological age, each Brain Age index and Brain Cognition as the regressors for fluid cognition. We found R^2 for these models at M=0.385 (SD = 0.042). 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 $\Delta R^2_{cognition}$ = 0.1183, statistically significant *p-value* at 0.05 in 24 out of 26 models).

For top-performing age/cognition-prediction models (e.g. Stacked All), the largest proportion of fluid cognition 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.

Discussion

To demonstrate the utility of Brain Age as a biomarker for fluid cognition, we investigated three essential 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 utility of Brain Age to capture fluid cognition above and beyond chronological age? The answer is also no. Third, how much does Brain Age miss the variation in the brain MRI that could explain fluid cognition? Brain Age and chronological age by themselves captured around 32% of the total variation in fluid cognition. But, around an additional 11% of the variation in fluid cognition could have been captured if we used the prediction models that directly predicted fluid cognition from brain MRI.

First, Brain Age itself did not add much more information to help us capture fluid cognition 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 the variation in fluid cognition. Including Brain Age indices only added around 1.6% at best. We believe the small unique effects of Brain Age were inevitable because, by design, Brain Age is tightly close to chronological age. Therefore, chronological age and Brain Age captured mostly a similar variation in fluid cognition.

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 fluid cognition. 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 fluid cognition 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 fluid cognition. 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 utility 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 on 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 fluid cognition (*Butler et al., 2021*). In our study, this can be seen in the small common effects between Corrected Brain Age Gap and chronological age in
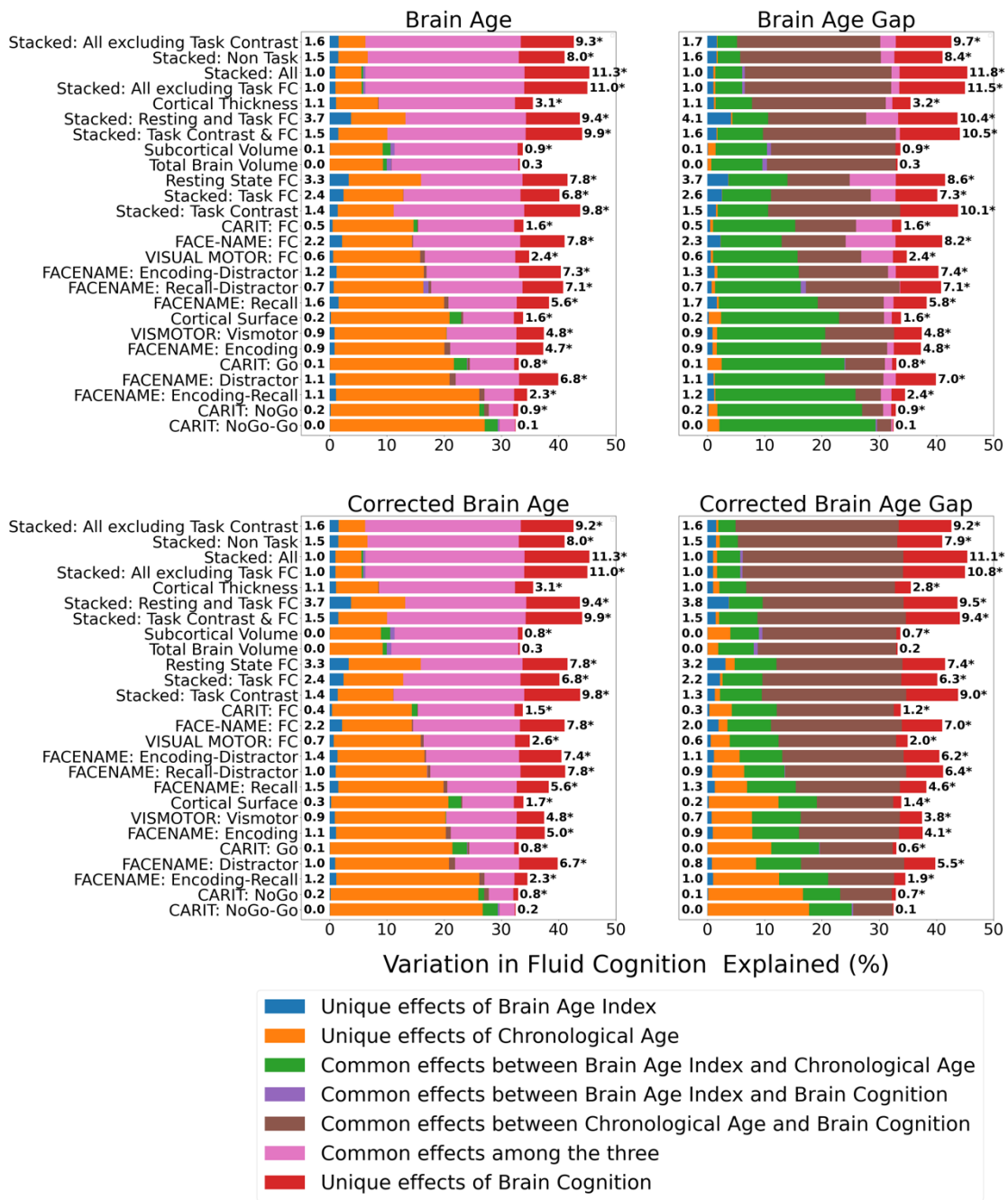
Figure 7. Commonality analysis of multiple regressions, having chronological age and each Brain Age index and Brain Cognition as the regressors for capturing fluid cognition. The numbers to the left of the figures represent the unique effects of Brain Age Index in %, and the numbers to the right of the figures represent the unique effects of Brain Cognition in %. * represents the statistical significance of the unique effects of Brain Cognition at $p<0.05$.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. Commonality analysis of multiple regressions, having chronological age, a quadratic term for chronological age and each Brain Age index as the regressors for fluid cognition.

Figure supplement 2. Commonality analysis of multiple regressions, having chronological age, a quadratic term for chronological age and each Brain Age index and Brain Cognition as the regressors for fluid cognition.

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 and 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 fluid cognition. 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.

The small effects of the Corrected Brain Age Gap in explaining fluid cognition of aging individuals found here are consistent with studies in older adults (*Cole, 2020*) and younger populations (*Butler et al., 2021*; *Jirsaraie et al., 2023b*). *Cole, 2020* studied the utility of Brain Age on cognitive functioning of large samples (n>17,000) of older adults, aged 45–80 years, from the UK Biobank (*Sudlow et al., 2015*). He constructed age-prediction models using LASSO, a similar penalised regression to ours and applied the same age-dependency adjustment to ours. *Cole, 2020* then conducted a multiple regression explaining cognitive functioning from Corrected Brain Age Gap while controlling for chronological age and other potential confounds. He found Corrected Brain Age Gap to be significantly related to performance in four out of six cognitive measures, and among those significant relationships, the effect sizes were small with a maximum of partial eta-squared at.0059. Similarly, *Jirsaraie et al., 2023a* studied the utility of Brain Age on cognitive functioning of youths aged 8–22 years old from the Human Connectome Project in Development (*Somerville et al., 2018*) and Preschool Depression Study (*Luby, 2010*). They built age-prediction models using gradient tree boosting (GTB) and deep-learning brain network (DBN) and adjusted the age dependency of Brain Age Gap using *Smith et al., 2019* method. Using multiple regressions, *Jirsaraie et al., 2023b* found weak effects of the adjusted Brain Age Gap on cognitive functioning across five cognitive tasks, five age-prediction models and the two datasets (mean of standardised regression coefficient = –0.09, see their Table S7). Next, *Butler et al., 2021* studied the utility of Brain Age on cognitive functioning of another group of youths aged 8–22 years old from the Philadelphia Neurodevelopmental Cohort (PNC) (*Satterthwaite et al., 2016*). Here, they used Elastic Net to build age-prediction models and applied another age-dependency adjustment method, proposed by *Beheshti et al., 2019*. Similar to the aforementioned results, *Butler et al., 2021* found a weak, statistically non-significant correlation between the adjusted Brain Age Gap and cognitive functioning at r=-0.01, p=0.71. Accordingly, the utility of Brain Age in explaining cognitive functioning beyond chronological age appears to be weak across age groups, different predictive modelling algorithms and age-dependency adjustments.

Second, the predictive performance of age-prediction models did not correspond to the utility of Brain Age in capturing fluid cognition over and above chronological age. For instance, while the best-performing age-prediction model was 'Stacked: All excluding Task Contrast' (R^2=0.775), the unique effects of Brain Age indices from this model in the two-regressor multiple regressions (i.e. with a Brain Age index and chronological age as regressor) were weak ($\Delta R^2_{\text{Brain Age index}}$ ≤0.0048) 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 ($\Delta R^2_{\text{Brain Age index}}$ ≤0.0135, p<0.05) that had a poorer performance in predicting chronological age (R^2=0.204). 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 fluid cognition. This discrepancy between the predictive performance of age-prediction models and the utility of Brain Age indices as a biomarker is consistent with recent findings (for review, see *Jirsaraie et al., 2023a*), both in the context of cognitive functioning (*Jirsaraie et al., 2023b*) and neurological/psychological disorders (*Bashyam et al., 2020*; *Rokicki et al., 2021*). For instance, combining different MRI modalities into the prediction models, similar to our stacked models, often leads to the highest performance of age-prediction models, but does not likely explain the highest variance across different phenotypes, including cognitive functioning and beyond (*Jirsaraie et al., 2023a*).

Third, by introducing Brain Cognition, we showed the extent to which Brain Age indices were not able to capture the variation in fluid cognition that is related to brain MRI. More specifically, using

Brain Cognition allowed us to gauge the variation in fluid cognition that is related to the brain MRI, and thereby, to estimate the upper limit of what Brain Age can do. Moreover, by examining what was captured by Brain Cognition, over and above Brain Age and chronological age via the unique effects of Brain Cognition, we were able to quantify the amount of co-variation between brain MRI and fluid cognition that was missed by Brain Age.

From our results, Brain Cognition, especially from certain cognition-prediction models such as the stacked models, has relatively good predictive performance, consistent with previous studies (*Dubois et al., 2018*; *Pat et al., 2022*; *Rasero et al., 2021*; *Sripada et al., 2020*; *Tetereva et al., 2022*; for review, see *Vieira et al., 2022*). We then examined Brain Cognition using commonality analyses (*Nimon et al., 2008*) in multiple regression models having a Brain Age index, chronological age and Brain Cognition as regressors to explain fluid cognition. Similar to Brain Age indices, Brain Cognition exhibited large common effects with chronological age. But more importantly, unlike Brain Age indices, Brain Cognition showed large unique effects, up to around 11%. As explained above, the unique effects of Brain Cognition indicated the amount of co-variation between brain MRI and fluid cognition that was missed by a Brain Age index and chronological age. This missing amount was relatively high, considering that Brain Age and chronological age together explained around 32% of the total variation in fluid cognition. Accordingly, if a Brain Age index was used as a biomarker along with chronological age, we would have missed an opportunity to improve the performance of the model by around one-third of the variation explained.

There are several potential limitations of this study. First, we conducted an investigation relying only on one dataset, the Human Connectome Project in Aging (HCP-A; *Bookheimer et al., 2019*). While HCP-A used state-of-the-art MRI methodologies, covered a wide age range from 36 to 100 years old and used several task-fMRI from different tasks that are harder to find in other bigger databases (e.g. UK Biobank from *Sudlow et al., 2015*), several characteristics of HCP-A might limit the generalisability of our findings. For instance, the tasks used in task-based fMRI in HCP-A are not used widely in clinical settings (*Horien et al., 2021*). This might make it challenging to translate the approaches used here. Similarly, HCP-A also excluded participants with neurological conditions, possibly making their participants not representative of the general population. Next, while HCP-A's sample size is not small (n=725 and 504 people, before and after exclusion, respectively), other datasets provide a much larger sample size (*Horien et al., 2021*). Similarly, HCP-A does not include younger populations. But as mentioned above, a study with a larger sample in older adults (*Cole, 2020*) and studies in younger populations (8–22 years old; *Butler et al., 2021*; *Jirsaraie et al., 2023b*) also found small effects of the adjusted Brain Age Gap in explaining cognitive functioning. And the disagreement between the predictive performance of age-prediction models and the utility of Brain Age found here is largely in line with the findings across different phenotypes seen in a recent systematic review (*Jirsaraie et al., 2023a*).

There is a notable difference between studies investigating the utility of Brain Age in explaining cognitive functioning, including ours and others (e.g. *Butler et al., 2021*; *Cole, 2020*; *Cole, 2020*; *Jirsaraie et al., 2023b*) and those explaining neurological/psychological disorders (e.g. *Bashyam et al., 2020*; *Rokicki et al., 2021*). We consider the former as a normative type of study and the latter as a case-control type of study (*Insel et al., 2010*; *Marquand et al., 2016*). Those case-control Brain Age studies focusing on neurological/psychological disorders often build age-prediction models from MRI data of largely healthy participants (e.g. controls in a case-control design or large samples in a population-based design), apply the built age-prediction models to participants without vs. with neurological/psychological disorders and compare Brain Age indices between the two groups. On the one hand, this means that case-control studies treat Brain Age as a method to detect anomalies in the neurological/psychological group (*Hahn et al., 2021*). On the other hand, this also means that case-control studies have to ignore under-fitted models when applied prediction models built from largely healthy participants to participants with neurological/psychological disorders (i.e. Brain Age may predict chronological age well for the controls, but not for those with a disorder). On the contrary, our study and other normative studies focusing on cognitive functioning often build age-prediction models from MRI data of largely healthy participants and apply the built age-prediction models to participants who are also largely healthy. Accordingly, the age-prediction models for explaining cognitive functioning in normative studies, while not allowing us to detect group-level anomalies, do not suffer from being under-fitted. This unfortunately might limit the generalisability of our study into just

the normative type of study. Future work is still needed to test the utility of brain age in the case-control case.

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. *Butler et al., 2021* recently highlighted this point, "These results indicate that the association between cognition and the BAG [Brain Age Gap] are driven by the association between age and cognitive performance. As such, it is critical that readers of past literature note whether or not age was controlled for when testing for effects on the BAG, as this has not always been common practice (*p.* 4097)." Similar to previous recommendations (*Butler et al., 2021*; *Le et al., 2018*), we suggest future work should account for the relationship between Brain Age and chronological age, either using Corrected Brain Age Gap (or other similar adjustments) or, better, examining unique effects of Brain Age indices after controlling for chronological age through commonality analyses. Note we prefer using the commonality analysis as it can decompose variance of the phenotype of interest into unique and common effects, allowing us to understand the shared variance between chronological age and Brain Age indices (*Ray-Mukherjee et al., 2014*). In our case, Brain Age indices had the same unique effects regardless of the level of common effects they had with chronological age (e.g. Brain Age vs. Corrected Brain Age Gap from stacked models). In the case of fluid cognition, the unique effects might be too small to be clinically meaningful as shown here and previously (*Butler et al., 2021*; *Cole, 2020*; *Jirsaraie et al., 2023b*).

Next, researchers should not select age-prediction models based solely on age-prediction performance. Instead, researchers could select age-prediction models that explained phenotypes of interest the best. Here, we selected age-prediction models based on a set of features (i.e., modalities) of brain MRI. This strategy was found effective not only for fluid cognition as we demonstrated here, but also for neurological and psychological disorders as shown elsewhere (*Jirsaraie et al., 2023a*; *Rokicki et al., 2021*). *Rokicki et al., 2021*, for instance, found that, while integrating across MRI modalities led to age-prediction models with the highest age-prediction performance, using only T1 structural MRI gave age-prediction models that were better at classifying Alzheimer's disease. Similarly, using only cerebral blood flow gave age-prediction models that were better at classifying mild/subjective cognitive impairment, schizophrenia and bipolar disorder.

As opposed to selecting age-prediction models based on a set of features, researchers could also select age-prediction models based on modelling methods. For instance, *Jirsaraie et al., 2023a* compared gradient tree boosting (GTB) and deep-learning brain network (DBN) algorithms in building age-prediction models. They found GTB to have higher age-prediction performance but DBN to have better utility in explaining cognitive functioning. In this case, an algorithm with better utility (e.g. DBN) should be used for explaining a phenotype of interest. *Bashyam et al., 2020* made a similar observation, though for a contradictory conclusion, see *Hahn et al., 2021*. Bashyam and colleagues built different DBN-based age-prediction models, varying in age-prediction performance. The DBN models with a higher number of epochs corresponded to higher age-prediction performance. However, DBN-based age-prediction models with a moderate (as opposed to higher or lower) number of epochs were better at classifying Alzheimer's disease, mild cognitive impairment and schizophrenia. In this case, a model from the same algorithm with better utility (e.g. those DBN with a moderate epoch number) should be used for explaining a phenotype of interest. In any case, this calls for a change in research practice, as recently pointed out by *Jirsaraie et al., 2023a*, "Despite mounting evidence, there is a persisting assumption across several studies that the most accurate brain age models will have the most potential for detecting differences in a given phenotype of interest". Future neuroimaging research should aim to build age-prediction models that are not necessarily good at predicting age, but at capturing phenotypes of interest.

Finally, researchers should test how much Brain Age miss the variation in the brain MRI that could explain fluid cognition or other phenotypes of interest. As demonstrated here, one straightforward method is to build a prediction model using a phenotype of interest as the target (e.g. fluid cognition) and incorporate the predicted value of this model (e.g. Brain Cognition), along with Brain Age and chronological age, into a multiple regression for commonality analyses. The unique effect of this predicted value will inform the missing variation in the brain MRI from Brain Age. If this unique effect is large then researchers might need to reconsider whether using Brain Age is appropriate for a particular phenotype of interest.

Altogether, we examined the utility of Brain Age as a biomarker for fluid cognition. Here are the three conclusions. First, Brain Age failed to add substantially more information over and above chronological age. Second, a higher ability to predict chronological age did not correspond to a higher utility to capture fluid cognition. Third, Brain Age missed up to around one-third of the variation in fluid cognition that could have been explained by brain MRI. Yet, given our focus on fluid cognition, future empirical research is needed to test the utility of Brain Age on other phenotypes, especially when Brain Age is used for anomaly detection in case-control studies (e.g. *Bashyam et al., 2020*; *Rokicki et al., 2021*). We hope that future studies may consider applying our approach (i.e. using the commonality analysis that includes predicted values from a model that directly predicts the phenotype of interest) to test the utility of Brain Age as a biomarker for other phenotypes.

Materials and methods

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. Note there were four individuals who were over 90 years old. HCP-A coded the age of these 90+individuals as 100 years old to reduce the leakage of their personal health information (see https://groups.google.com/a/humanconnectome.org/g/hcp-users/c/esZTVCRuxwE/m/xx4PLYMlCQAJ). For ethical procedures including informed consent and the demographics of the participants, please see *Bookheimer et al., 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.

Sets of features 1-10: Task fMRI contrast (Task Contrast)

Task contrasts reflect fMRI activation relevant to events in each task. *Bookheimer et al., 2019* provided detailed information about the fMRI in HCP-A. Here, we focused on the pre-processed task fMRI Connectivity Informatics Technology Initiative (CIFTI) files with a suffix, '_PA_Atlas_MSMAll_hp0_clean.dtseries.nii'. These CIFTI files encompassed both the cortical mesh surface and subcortical volume (*Glasser et al., 2013*). Collected using the posterior-to-anterior (PA) phase, these files were aligned using MSMALL (*Glasser et al., 2016*; *Robinson et al., 2018*), linear detrended (see https://groups.google.com/a/humanconnectome.org/g/hcp-users/c/ZLJc092h980/m/GiihzQAUAwAJ) and cleaned from potential artifacts using ICA-FIX (*Glasser et al., 2016*).

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 kept FSL's default high pass cutoff at 200 s (i.e. 0.005 Hz). 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, whose number 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*). We used the same CIFTI file '_PA_Atlas_MSMAll_hp0_clean.dtseries.nii.' as the task contrasts. 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*). We computed these regressors on FSL, and regressed them in nilearn (*Abraham et al., 2014*). Following previous work on task FC (*Elliott et al., 2019*), we applied a highpass at.008 Hz. For parcellation, we used 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*; *Sripada et al., 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 2 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. highpass filter, 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.

Fluid cognition

We measured fluid cognition via the NIH Toolbox (*Weintraub et al., 2014*), using the 'fluidcogcomp_unadj' variable. Fluid cognition summarises scores from five tests assessed outside of the MRI:
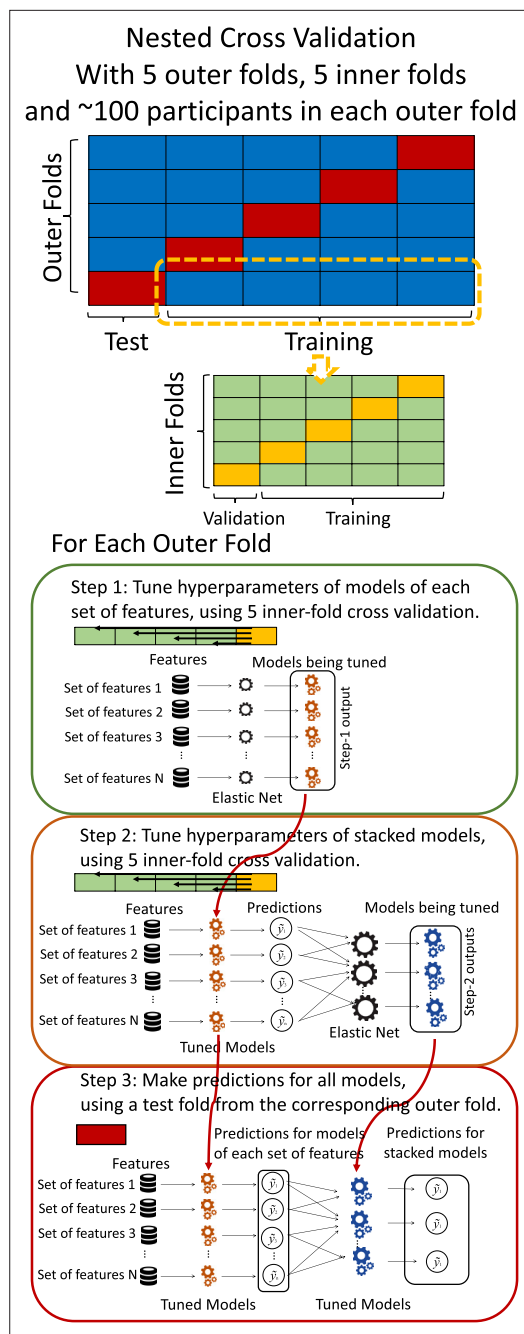
Figure 8. Diagram of the nested cross-validation used for creating predictions for models of each set of features as well as predictions for stacked models.

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 fluid cognition as the target and standardised brain MRI as the features (*Denissen et al., 2022*). We used nested cross-validation (CV) to build these prediction models (see *Figure 8*). We first split the data into five outer folds, leaving each outer fold with around 100 participants. This number of participants in each fold is to ensure the stability of the test performance across folds. In each outer-fold CV loop, one of the outer folds was treated as an outer-fold test set, and the rest was treated as an outer-fold training set. Ultimately, looping through the nested CV resulted in (a) prediction models from each of the 18 sets of features as well as (b) prediction models that drew information across different combinations of the 18 separate sets, known as '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, there were 26 prediction models in total for both Brain Age and Brain Cognition.

To create these 26 prediction models, we applied three steps for each outer-fold loop. The first step aimed at tuning prediction models for each of 18 sets of features. This step only involved the outer-fold training set and did not involve the outer-fold test set. Here, we divided the outer-fold training set into five inner folds and applied inner-fold CV to tune hyperparameters with grid search. Specifically, in each inner-fold CV, one of the inner folds was treated as an inner-fold validation set, and the rest was treated as an inner-fold training set. Within each inner-fold CV loop, we used the inner-fold training set to estimate parameters of the prediction model with a particular set

of hyperparameters and applied the estimated model to the inner-fold validation set. After looping through the inner-fold CV, we, then, chose the prediction models that led to the highest performance, reflected by coefficient of determination (R^2), on average across the inner-fold validation sets. This led to 18 tuned models, one for each of the 18 sets of features, for each outer fold.

The second step aimed at tuning stacked models. Same as the first step, the second step only involved the outer-fold training set and did not involve the outer-fold test set. Here, using the same outer-fold training set as the first step, we applied tuned models, created from the first step, one

from each of the 18 sets of features, resulting in 18 predicted values for each participant. We, then, re-divided this outer-fold training set into new five inner folds. In each inner fold, we treated different combinations of the 18 predicted values from separate sets of features as features to predict the targets in separate 'stacked' models. Same as the first step, in each inner-fold CV loop, we treated one out of five inner folds as an inner-fold validation set, and the rest as an inner-fold training set. Also as in the first step, we used the inner-fold training set to estimate parameters of the prediction model with a particular set of hyperparameters from our grid. We tuned the hyperparameters of stacked models using grid search by selecting the models with the highest R^2 on average across the inner-fold validation sets. This led to eight tuned stacked models.

The third step aimed at testing the predictive performance of the 18 tuned prediction models from each of the set of features, built from the first step, and eight tuned stacked models, built from the second step. Unlike the first two steps, here we applied the already tuned models to the outer-fold test set. We started by applying the 18 tuned prediction models from each of the sets of features to each observation in the outer-fold test set, resulting in 18 predicted values. We then applied the tuned stacked models to these predicted values from separate sets of features, resulting in eight predicted values.

To demonstrate the predictive performance, we assessed the similarity between the observed values and the predicted values of each model across outer-fold 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. $R^2 = 1 -$ (sum of squares residuals/total sum of squares)) per a previous recommendation (*Poldrack et al., 2020*). We considered the predicted values from the outer-fold test sets of models predicting age or fluid cognition, as Brain Age and Brain Cognition, respectively.

We controlled for the potential influences of biological sex on the brain features by first residual-ising biological sex from brain features in each outer-fold training set. We then applied the regression of this residualisation to the corresponding outer-fold test set. We also standardised the brain features in each outer-fold training set and then used the mean and standard deviation of this outer-fold training set to standardise the outer-fold test set. All of the standardisation was done prior to fitting the prediction models.

For the machine learning algorithm, we used Elastic Net (*Zou and Hastie, 2005*). Elastic Net is a general form of penalised regressions (including Lasso and Ridge regression), allowing us to simultaneously draw information across different brain indices to predict one target variable. Penalised regressions are commonly used for building age-prediction models (*Jirsaraie et al., 2023a*). Previously we showed that the performance of Elastic Net in predicting cognitive abilities is on par, if not better than, many non-linear and more-complicated algorithms (*Pat et al., 2023*; *Tetereva et al., 2022*). Moreover, Elastic Net coefficients are readily explainable, allowing us the ability to explain how our age-prediction and cognition-prediction models made the prediction from each brain feature (*Molnar, 2019*; *Pat et al., 2023*) (see below).

Elastic Net simultaneously minimises the weighted sum of the features' coefficients. The degree of penalty to the sum of the feature's coefficients is determined by a shrinkage hyperparameter 'α': the greater the α, the more the coefficients shrink, and the more regularised the model becomes. Elastic Net also includes another hyperparameter, ℓ ratio', which determines the degree to which the sum of either the squared (known as 'Ridge'; ℓ ratio = 0) or absolute (known as 'Lasso'; ℓ ratio = 1) coefficients is penalised (*Zou and Hastie, 2005*). The objective function of Elastic Net as implemented by sklearn (*Pedregosa et al., 2011*) is defined as:

$$\mathrm{argmin}_\beta \left(\frac{||y - X||_2^2}{2 \times n_{samples}} + \alpha \times \ell_1\_ratio \times ||\beta||_1 + 0.5 \times \alpha \times \left(1 - \ell_1\_ratio\right) \times ||w\beta||_2^2 \right), \qquad (1)$$

where X is the features, y is the target, and β is the coefficient. In our grid search, we tuned two Elastic Net hyperparameters: α using 70 numbers in log space, ranging from.1 and 100, and ℓ -ratio using 25 numbers in linear space, ranging from 0 and 1.

To understand how Elastic Net made a prediction based on different brain features, we examined the coefficients of the tuned model. Elastic Net coefficients can be considered as feature importance, such that more positive Elastic Net coefficients lead to more positive predicted values and, similarly, more negative Elastic Net coefficients lead to more negative predicted values (*Molnar, 2019*; *Pat et al., 2023*). While the magnitude of Elastic Net coefficients is regularised (thus making it difficult for

us to interpret the magnitude itself directly), we could still indicate that a brain feature with a higher magnitude weights relatively stronger in making a prediction. Another benefit of Elastic Net as a penalised regression is that the coefficients are less susceptible to collinearity among features as they have already been regularised (**Dormann et al., 2013**; **Pat et al., 2022**).

Given that we used fivefold nested cross validation, different outer folds may have different degrees of 'α' and 'l₁ ratio', making the final coefficients from different folds to be different. For instance, for certain sets of features, penalisation may not play a big part (i.e., higher or lower 'α' leads to similar predictive performance), resulting in different 'α' for different folds. To remedy this in the visualisation of Elastic Net feature importance, we refitted the Elastic Net model to the full dataset without splitting them into five folds and visualised the coefficients on brain images using Brainspace (**Vos de Wael et al., 2020**) and Nilern (**Abraham et al., 2014**) packages. Note, unlike other sets of features, Task FC and Rest FC were modelled after data reduction via PCA. Thus, for Task FC and Rest FC, we, first, multiplied the absolute PCA scores (extracted from the 'components_' attribute of 'sklearn. decomposition.PCA') with Elastic Net coefficients and, then, summed the multiplied values across the 75 components, leaving 71,631 ROI-pair indices.

To demonstrate the stability of feature importance across outer folds, we examined the rank stability of feature importance using Spearman's ρ. Specifically, we correlated the feature importance between two prediction models of the same features, used in two different outer-fold test sets. Given that there were five outer-fold test sets, we computed 10 Spearman's ρ for each prediction model of the same features.

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 outer-fold 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:

$$Brain\,Age\,Gap_i = Brain\,Age_i - chronological\,age_i, \tag{2}$$

where i is the individual. Next, to reduce the dependency on chronological age (**Butler et al., 2021**; **de Lange and Cole, 2020**; **Le et al., 2018**), we applied a method described in **de Lange and Cole, 2020**, which was implemented elsewhere (**Cole et al., 2020**; **Cumplido-Mayoral et al., 2023**; **Denissen et al., 2022**):

In each outer-fold training set:

$$Brain\,Age_i = \beta_0 + \beta_1 chronological\,age_i + \varepsilon_i, \tag{3}$$

Then   in   the   corresponding   outer-fold   test   set:

$$Corrected\,Brain\,Age_i = (Brain\,Age_i - \beta_0)/\beta_1, \tag{4}$$

That is, we first fit a regression line predicting the Brain Age from a chronological age in each outer-fold training set. We then used the slope (β_1) and intercept (β_0) of this regression line to adjust Brain Age in the corresponding outer-fold test set, resulting in Corrected Brain Age. Note **de Lange and Cole, 2020** called this Corrected Brain Age, 'Corrected Predicted Age', while **Butler et al., 2021** called it 'Revised Predicted Age'.

Lastly, we computed Corrected Brain Age Gap by subtracting the chronological age from the Corrected Brain Age (**Butler et al., 2021**; **Cole et al., 2020**; **de Lange and Cole, 2020**; **Denissen et al., 2022**):

$$Corrected\,Brain\,Age\,Gap = Corrected\,Brain\,Age - chronological\,age, \tag{5}$$

Note **Cole, 2020** called Corrected Brain Age Gap, 'brain-predicted age difference (brain-PAD)', while **Butler et al., 2021** called this index, 'Revised Brain Age Gap'.

The utility of Brain Age indices to capture fluid cognition

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

Simple regression: using each Brain Age index to explain fluid cognition

Here using simple regression, we simply had each Brain Age index as the sole regressor for fluid cognition:

$$Fluid\ Cognition_i = \beta_0 + \beta_1\ Brain\ Age\ Index_{i,j} + \varepsilon_i, \tag{6}$$

where j 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 fluid cognition 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 fluid cognition.

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

$$Fluid\ Cognition_i = \beta_0 + \beta_1\ Brain\ Cognition_{i,j} \tag{7}$$

This allows us to compare the utility of Brain Age Indices vs. Brain Cognition as a sole regressor for predicting fluid cognition.

Multiple regression: using chronological age and each Brain Age index to explain fluid cognition

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

$$Fluid\ Cognition_i = \beta_0 + \beta_1\ Chonological\ Age_i + \beta_2\ Brain\ Age\ Index_{i,j} + \varepsilon_i \tag{8}$$

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 fluid cognition 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:

$$UniqueEffect_{chronologicalage} = \Delta R^2_{chronologicalage} = R^2_{chronologicalage,\ Brain\ Age\ Index} - R^2_{Brain\ Age\ index}$$
$$UniqueEffect_{BrainAgeindex} = \Delta R^2_{BrainAgeindex} = R^2_{chronologicalage,\ Brain\ Age\ Index} - R^2_{chronologicalage}, \tag{9}$$

We determined the statistical significance of ΔR^2 by:

$$F_{Change} = \frac{(N - k_2 - 1)\,\Delta R^2}{k_{change}\,(1 - R^2_2)}, \tag{10}$$

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:

$$Common\ Effect_{chronological\ age,\ Brain\ Age\ index} = R^2_{chronological\ age,\ Brain\ Age\ index} - \Delta R^2_{chronological\ age}$$
$$-\Delta R^2_{Brain\ Age\ index}, \tag{11}$$

These common effects indicate the extent to which variation in fluid cognition 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.

Multiple regression: using chronological age, each Brain Age index and Brain Cognition to explain fluid cognition

Similar to the above multiple regression model, we had chronological age, each Brain Age index and Brain Cognition as the regressors for fluid cognition:

$$Fluid\ Cognition_i = \beta_0 + \beta_1 Chronological\ Age_i + \beta_2 Brain\ Age\ Index_{i,j} + \beta_3 Brain\ cognition_i + \varepsilon_i, \tag{12}$$

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 fluid cognition. We calculated the commonality analysis as follows *Nimon et al., 2017* (Table 2):

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

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

$$UniqueEffect_{Brain\ cognition} = \Delta R^2_{Brain\ cognition} = R^2_{chronological\ age,\ Brain\ Age\ index,\ Brain\ Cognition} - R^2_{chronological\ Age\ index,\ Brain\ Age\ Index}$$

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

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

$$Common\ Effect_{Brain\ Age\ index,\ Brain\ cognition} = R^2_{chronological\ age,\ Brain\ Age\ Index} + R^2_{chronological\ age,\ Braincognition} - R^2_{chronological\ age} - R^2_{chronological\ a}$$

$$Common\ Effect_{chronological\ age,\ Brain\ Age\ Index} = R^2_{chronological\ age} + R^2_{Brain\ Age\ index} - R^2_{Brain\ cognition} - R^2_{chronological\ age,\ Brain\ Age\ index} - R^2_{chronologica}$$

$$\tag{13}$$

Note to ensure that the commonality analysis results were robust against multicollinearity (*Ray-Mukherjee et al., 2014*), we also repeated the same commonality analyses done here on Ridge regression, as opposed to multiple regression. Ridge regression is a method designed to deal with multicollinearity (*Dormann et al., 2013*). See *Figure 6—figure supplement 1* for the Ridge regression with chronological age and each Brain Age index as regressors and *Figure 6—figure supplement 2* for the Ridge regression with chronological age, each Brain Age and Brain Cognition index as regressors. Briefly, the results from commonality analyses applied to Ridge regressions are closely matched with our results done using multiple regression.

Similarly, to ensure that we were able to capture the non-linear pattern of chronological age in explaining fluid cognition, we added a quadratic term of chronological age to our multiple-regression models in the commonality analyses. See *Figure 7—figure supplement 1* for the multiple regression with chronological age, square chronological age and each Brain Age index as regressors and *Figure 7—figure supplement 2* for the multiple regression with chronological age, square chronological age, each Brain Age index and Brain Cognition as regressors. Briefly, adding the quadratic term for chronological age did not change the pattern of the results of the commonality analyses.

Acknowledgements

Additional information

Author contributions

Alina Tetereva, Data curation, Software, Formal analysis, Validation, Visualization, Methodology, Writing - original draft; Narun Pat, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Alina Tetereva ⓘ http://orcid.org/0000-0003-2077-628X
Narun Pat ⓘ http://orcid.org/0000-0003-1459-5255

Ethics

We used an openly available dataset, "The Lifespan Human Connectome Project in Aging". For ethical procedures including informed consent, please see Bookheimer and colleagues' (2019). Bookheimer SY, Salat DH, Terpstra M, Ances BM, Barch DM, Buckner RL, Burgess GC, Curtiss SW, Diaz-Santos M, Elam JS, Fischl B, Greve DN, Hagy HA, Harms MP, Hatch OM, Hedden T, Hodge C, Japardi KC, Kuhn TP, Ly TK, Smith SM, Somerville LH, Uğurbil K, van der Kouwe A, Van Essen D, Woods RP, 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.

Peer review material

Reviewer #1 (Public Review): https://doi.org/10.7554/eLife.87297.4.sa1
Author response https://doi.org/10.7554/eLife.87297.4.sa2

Additional files

Supplementary files
• MDAR checklist

Data availability

We used the openly available data provided by the Human Connectome Project in Aging. Please see here for data access and download instructions. The shell and Python scripts used in the analyses are made available on GitHub (copy archived at *Tetereva, 2024*).

References

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G. 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics* **8**:14. DOI: https://doi.org/10.3389/fninf.2014.00014, PMID: 24600388

Ances BM, Liang CL, Leontiev O, Perthen JE, Fleisher AS, Lansing AE, Buxton RB. 2009. Effects of aging on cerebral blood flow, oxygen metabolism, and blood oxygenation level dependent responses to visual stimulation. *Human Brain Mapping* **30**:1120–1132. DOI: https://doi.org/10.1002/hbm.20574, PMID: 18465743

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. DOI: https://doi.org/10.1016/j.ebiom.2021.103600, PMID: 34614461

Bashyam VM, Erus G, Doshi J, Habes M, Nasrallah IM, Truelove-Hill M, Srinivasan D, Mamourian L, Pomponio R, Fan Y, Launer LJ, Masters CL, Maruff P, Zhuo C, Völzke H, Johnson SC, Fripp J, Koutsouleris N, Satterthwaite TD, Wolf D, et al. 2020. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**:2312–2324. DOI: https://doi.org/10.1093/brain/awaa160

Beheshti I, Nugent S, Potvin O, Duchesne S. 2019. Bias-adjustment in neuroimaging-based brain age frameworks: a robust scheme. *NeuroImage. Clinical* **24**:102063. DOI: https://doi.org/10.1016/j.nicl.2019.102063, PMID: 31795063

Bookheimer SY, Salat DH, Terpstra M, Ances BM, Barch DM, Buckner RL, Burgess GC, Curtiss SW, Diaz-Santos M, Elam JS, Fischl B, Greve DN, Hagy HA, Harms MP, Hatch OM, Hedden T, Hodge C, Japardi KC, Kuhn TP, Ly TK, et al. 2019. The lifespan human connectome project in aging: an overview. *NeuroImage* **185**:335–348. DOI: https://doi.org/10.1016/j.neuroimage.2018.10.009, PMID: 30332613

Boyle R, Jollans L, Rueda-Delgado LM, Rizzo R, Yener GG, McMorrow JP, Knight SP, Carey D, Robertson IH, Emek-Savaş DD, Stern Y, Kenny RA, Whelan R. 2021. Brain-predicted age difference score is related to specific cognitive functions: a multi-site replication analysis. *Brain Imaging and Behavior* **15**:327–345. DOI: https://doi.org/10.1007/s11682-020-00260-3, PMID: 32141032

Butler ER, Chen A, Ramadan R, Le TT, Ruparel K, Moore TM, Satterthwaite TD, Zhang F, Shou H, Gur RC, Nichols TE, Shinohara RT. 2021. Pitfalls in brain age analyses. *Human Brain Mapping* **42**:4092–4101. DOI: https://doi.org/10.1002/hbm.25533, PMID: 34190372

Cole JH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N, Corley J, Pattie A, Harris SE, Zhang Q, Wray NR, Redmond P, Marioni RE, Starr JM, Cox SR, Wardlaw JM, Sharp DJ, Deary IJ. 2018. Brain age predicts mortality. *Molecular Psychiatry* **23**:1385–1392. DOI: https://doi.org/10.1038/mp.2017.62, PMID: 28439103

Cole JH. 2020. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of Aging* **92**:34–42. DOI: https://doi.org/10.1016/j.neurobiolaging.2020.03.014, PMID: 32380363

Cole JH, Raffel J, Friede T, Eshaghi A, Brownlee WJ, Chard D, De Stefano N, Enzinger C, Pirpamer L, Filippi M, Gasperini C, Rocca MA, Rovira A, Ruggieri S, Sastre-Garriga J, Stromillo ML, Uitdehaag BMJ, Vrenken H, Barkhof F, Nicholas R, et al. 2020. Longitudinal assessment of multiple sclerosis with the brain-age paradigm. *Annals of Neurology* **88**:93–105. DOI: https://doi.org/10.1002/ana.25746, PMID: 32285956

Cumplido-Mayoral I, García-Prat M, Operto G, Falcon C, Shekari M, Cacciaglia R, Milà-Alomà M, Lorenzini L, Ingala S, Meije Wink A, Mutsaerts HJMM, Minguillón C, Fauria K, Molinuevo JL, Haller S, Chetelat G, Waldman A, Schwarz AJ, Barkhof F, Suridjan I, et al. 2023. Biological brain age prediction using machine learning on structural neuroimaging data: multi-cohort validation against biomarkers of alzheimer's disease and neurodegeneration stratified by sex. *eLife* **12**:e81067. DOI: https://doi.org/10.7554/eLife.81067, PMID: 37067031

de Lange A-MG, Cole JH. 2020. Commentary: correction procedures in brain-age prediction. *NeuroImage. Clinical* **26**:102229. DOI: https://doi.org/10.1016/j.nicl.2020.102229, PMID: 32120292

Denissen S, Engemann DA, De Cock A, Costers L, Baijot J, Laton J, Penner I-K, Grothe M, Kirsch M, D'hooghe MB, D'Haeseleer M, Dive D, De Mey J, Van Schependom J, Sima DM, Nagels G. 2022. Brain age as a surrogate marker for cognitive performance in multiple sclerosis. *European Journal of Neurology* **29**:3039–3049. DOI: https://doi.org/10.1111/ene.15473, PMID: 35737867

Destrieux C, Fischl B, Dale A, Halgren E. 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**:1–15. DOI: https://doi.org/10.1016/j.neuroimage.2010.06.010, PMID: 20547229

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S. 2013. Collinearity: a review of methods to deal with it and A simulation study evaluating their performance. *Ecography* **36**:27–46. DOI: https://doi.org/10.1111/j.1600-0587.2012.07348.x

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

Elliott ML, Knodt AR, Cooke M, Kim MJ, Melzer TR, Keenan R, Ireland D, Ramrakha S, Poulton R, Caspi A, Moffitt TE, Hariri AR. 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. DOI: https://doi.org/10.1016/j.neuroimage.2019.01.068, PMID: 30708106

Engemann DA, 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. DOI: https://doi.org/10.7554/eLife.54055, PMID: 32423528

Fair DA, Schlaggar BL, Cohen AL, Miezin FM, Dosenbach NUF, Wenger KK, Fox MD, Snyder AZ, Raichle ME, Petersen SE. 2007. A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. *NeuroImage* **35**:396–405. DOI: https://doi.org/10.1016/j.neuroimage.2006.11.051, PMID: 17239622

Fischl B, Salat DH, 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 AM. 2002. Whole brain segmentation. *Neuron* **33**:341–355. DOI: https://doi.org/10.1016/S0896-6273(02)00569-X

Fischl B. 2012. FreeSurfer. *NeuroImage* **62**:774–781. DOI: https://doi.org/10.1016/j.neuroimage.2012.01.021, PMID: 22248573

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**:883–892. DOI: https://doi.org/10.1016/j.neuroimage.2010.01.005, PMID: 20070949

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. DOI: https://doi.org/10.3389/fneur.2019.00789, PMID: 31474922

Frederick BN. 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 GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, Démonet J-F, Garibotto V, Giannakopoulos P, Gietl A, Hansson O, Herholz K, Jack CR Jr, Nobili F, Nordberg A, Snyder HM, Ten Kate M, Varrone A, Albanese E, Becker S, et al. 2017. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *The Lancet. Neurology* **16**:661–676. DOI: https://doi.org/10.1016/S1474-4422(17)30159-X, PMID: 28721928

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M, WU-Minn HCP Consortium. 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage* **80**:105–124. DOI: https://doi.org/10.1016/j.neuroimage.2013.04.127, PMID: 23668970

Glasser MF, Smith SM, Marcus DS, Andersson JLR, Auerbach EJ, Behrens TEJ, Coalson TS, Harms MP, Jenkinson M, Moeller S, Robinson EC, Sotiropoulos SN, Xu J, Yacoub E, Ugurbil K, Van Essen DC. 2016. The human connectome project's neuroimaging approach. *Nature Neuroscience* **19**:1175–1187. DOI: https://doi.org/10.1038/nn.4361, PMID: 27571196

Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. 2016. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex* **26**:288–303. DOI: https://doi.org/10.1093/cercor/bhu239, PMID: 25316338

Gratton C, Laumann TO, Nielsen AN, Greene DJ, Gordon EM, Gilmore AW, Nelson SM, Coalson RS, Snyder AZ, Schlaggar BL, Dosenbach NUF, Petersen SE. 2018. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* **98**:439–452.. DOI: https://doi.org/10.1016/j.neuron.2018.03.035, PMID: 29673485

Hahn T, Fisch L, Ernsting J, Winter NR, Leenings R, Sarink K, Emden D, Kircher T, Berger K, Dannlowski U. 2021. From "loose fitting" to high-performance, uncertainty-aware brain-age modelling. *Brain* **144**:e31. DOI: https://doi.org/10.1093/brain/awaa454, PMID: 33826702

Harms MP, Somerville LH, Ances BM, Andersson J, Barch DM, Bastiani M, Bookheimer SY, Brown TB, Buckner RL, Burgess GC, Coalson TS, Chappell MA, Dapretto M, Douaud G, Fischl B, Glasser MF, Greve DN, Hodge C, Jamison KW, Jbabdi S, et al. 2018. Extending the human connectome project across ages: imaging protocols for the lifespan development and aging projects. *NeuroImage* **183**:972–984. DOI: https://doi.org/10.1016/j.neuroimage.2018.09.060, PMID: 30261308

Horien C, Noble S, Greene AS, Lee K, Barron DS, Gao S, O'Connor D, Salehi M, Dadashkarimi J, Shen X, Lake EMR, Constable RT, Scheinost D. 2021. A hitchhiker's guide to working with large, open-source neuroimaging datasets. *Nature Human Behaviour* **5**:185–193. DOI: https://doi.org/10.1038/s41562-020-01005-4, PMID: 33288916

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry* **167**:748–751. DOI: https://doi.org/10.1176/appi.ajp.2010.09091379, PMID: 20595427

Jirsaraie RJ, Gorelik AJ, Gatavins MM, Engemann DA, Bogdan R, Barch DM, Sotiras A. 2023a. A systematic review of multimodal brain age studies: Uncovering A divergence between model accuracy and utility. *Patterns* **4**:100712. DOI: https://doi.org/10.1016/j.patter.2023.100712, PMID: 37123443

Jirsaraie RJ, Kaufmann T, Bashyam V, Erus G, Luby JL, Westlye LT, Davatzikos C, Barch DM, Sotiras A. 2023b. Benchmarking the generalizability of brain age models: challenges posed by scanner variance and prediction bias. *Human Brain Mapping* **44**:1118–1128. DOI: https://doi.org/10.1002/hbm.26144, PMID: 36346213

Jonsson BA, Bjornsdottir G, Thorgeirsson TE, Ellingsen LM, Walters GB, Gudbjartsson DF, Stefansson H, Stefansson K, Ulfarsson MO. 2019. Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications* **10**:5409. DOI: https://doi.org/10.1038/s41467-019-13163-9, PMID: 31776335

Le TT, Kuplicki RT, McKinney BA, Yeh H-W, Thompson WK, Paulus MP, Tulsa 1000 Investigators. 2018. A nonlinear simulation framework supports adjusting for age when analyzing BrainAGE. *Frontiers in Aging Neuroscience* **10**:317. DOI: https://doi.org/10.3389/fnagi.2018.00317, PMID: 30405393

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**:3143–3152. DOI: https://doi.org/10.1002/hbm.24588, PMID: 30924225

Liem F, Varoquaux G, Kynast J, Beyer F, Kharabian Masouleh S, Huntenburg JM, Lampe L, Rahim M, Abraham A, Craddock RC, Riedel-Heller S, Luck T, Loeffler M, Schroeter ML, Witte AV, Villringer A, Margulies DS. 2017. Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage* **148**:179–188. DOI: https://doi.org/10.1016/j.neuroimage.2016.11.005, PMID: 27890805

Luby JL. 2010. Preschool depression: the importance of identification of depression early in development. *Current Directions in Psychological Science* **19**:91–95. DOI: https://doi.org/10.1177/0963721410364493, PMID: 21969769

Marquand AF, Rezek I, Buitelaar J, Beckmann CF. 2016. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological Psychiatry* **80**:552–561. DOI: https://doi.org/10.1016/j.biopsych.2015.12.023, PMID: 26927419

Molnar C. 2019. *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*. Christoph Molnar.

Nimon K, Lewis M, Kane R, Haynes RM. 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**:457–466. DOI: https://doi.org/10.3758/brm.40.2.457, PMID: 18522056

Nimon K, Lewis M, Kane R, Haynes RM. 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**:2275. DOI: 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* **43**:5520–5542. DOI: https://doi.org/10.1002/hbm.26027, PMID: 35903877

Pat N, Wang Y, Bartonicek A, Candia J, Stringaris A. 2023. Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cerebral Cortex* **33**:2682–2703. DOI: 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**:2825–2830.

Poldrack RA, Huckins G, Varoquaux G. 2020. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* **77**:534–540. DOI: https://doi.org/10.1001/jamapsychiatry.2019.3671, PMID: 31774490

Rasero J, Sentis AI, Yeh FC, Verstynen T. 2021. Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLOS Computational Biology* **17**:e1008347. DOI: https://doi.org/10.1371/journal.pcbi.1008347, PMID: 33667224

Ray-Mukherjee J, Nimon K, Mukherjee S, Morris DW, 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**:320–328. DOI: https://doi.org/10.1111/2041-210X.12166

Richard G, Kolskår K, Sanders A-M, Kaufmann T, Petersen A, Doan NT, Monereo Sánchez J, Alnæs D, Ulrichsen KM, Dørum ES, Andreassen OA, Nordvik JE, Westlye LT. 2018. Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ* **6**:e5908. DOI: https://doi.org/10.7717/peerj.5908, PMID: 30533290

Robinson EC, Garcia K, Glasser MF, Chen Z, Coalson TS, Makropoulos A, Bozek J, Wright R, Schuh A, Webster M, Hutter J, Price A, Cordero Grande L, Hughes E, Tusor N, Bayly PV, Van Essen DC, Smith SM, Edwards AD, Hajnal J, et al. 2018. Multimodal surface matching with higher-order smoothness constraints. *NeuroImage* **167**:453–465. DOI: https://doi.org/10.1016/j.neuroimage.2017.10.037, PMID: 29100940

Rokicki J, Wolfers T, Nordhøy W, Tesli N, Quintana DS, Alnaes D, Richard G, de Lange A-MG, Lund MJ, Norbom L, Agartz I, Melle I, Naerland T, Selbaek G, Persson K, Nordvik JE, Schwarz E, Andreassen OA, Kaufmann T, Westlye LT. 2021. Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Human Brain Mapping* **42**:1714–1726. DOI: https://doi.org/10.1002/hbm.25323, PMID: 33340180

Salthouse TA. 2019. Trajectories of normal cognitive aging. *Psychology and Aging* **34**:17–24. DOI: https://doi.org/10.1037/pag0000288, PMID: 30211596

Satterthwaite TD, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA, Roalf DR, Hopson R, Prabhakaran K, Behr M, Qiu H, Mentch FD, Chiavacci R, Sleiman PMA, Gur RC, Hakonarson H, Gur RE. 2016. The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *NeuroImage* **124**:1115–1119. DOI: https://doi.org/10.1016/j.neuroimage.2015.03.056, PMID: 25840117

Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. 2019. Estimation of brain age delta from brain imaging. *NeuroImage* **200**:528–539. DOI: https://doi.org/10.1016/j.neuroimage.2019.06.017

Somerville LH, Bookheimer SY, Buckner RL, Burgess GC, Curtiss SW, Dapretto M, Elam JS, Gaffrey MS, Harms MP, Hodge C, Kandala S, Kastman EK, Nichols TE, Schlaggar BL, Smith SM, Thomas KM, Yacoub E, Van Essen DC, Barch DM. 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. DOI: https://doi.org/10.1016/j.neuroimage.2018.08.050, PMID: 30142446

Sperling RA, Bates JF, Cocchiarella AJ, Schacter DL, Rosen BR, Albert MS. 2001. Encoding novel face-name associations: a functional MRI study. *Human Brain Mapping* **14**:129–139. DOI: https://doi.org/10.1002/hbm.1047, PMID: 11559958

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**:1900. DOI: https://doi.org/10.1038/s41598-018-38406-5, PMID: 30760808

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**:3186–3197. DOI: https://doi.org/10.1002/hbm.25007, PMID: 32364670

Stigler SM. 1997. Regression towards the mean, historically considered. *Statistical Methods in Medical Research* **6**:103–114. DOI: https://doi.org/10.1177/096228029700600202, PMID: 9261910

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine* **12**:e1001779. DOI: https://doi.org/10.1371/journal.pmed.1001779, PMID: 25826379

Tetereva A, Li J, Deng JD, 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. DOI: https://doi.org/10.1016/j.neuroimage.2022.119588, PMID: 36057404

Tetereva A. 2024. HCP-aging Commonality. swh:1:rev:f19eab409ad0cb56c234d068733b2d6116f6d422. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:3fba14abb1c13debd408f0bbf998fc728c8e1121;origin=https://github.com/HAM-lab-Otago-University/HCP-Aging_commonality;visit=swh:1:snp:dda558f0751e6434c45a6c15f95287c88097d5f1;anchor=swh:1:rev:f19eab409ad0cb56c234d068733b2d6116f6d422

Vieira BH, Pamplona GSP, Fachinello K, Silva AK, Foss MP, Salmon CEG. 2022. On the prediction of human intelligence from neuroimaging: a systematic review of methods and reporting. *Intelligence* **93**:101654. DOI: https://doi.org/10.1016/j.intell.2022.101654

Vos de Wael R, Benkarim O, Paquola C, Lariviere S, Royer J, Tavakol S, Xu T, Hong SJ, Langs G, Valk S, Misic B, Milham M, Margulies D, Smallwood J, Bernhardt BC. 2020. BrainSpace: A toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology* **3**:020-0794-7. DOI: https://doi.org/10.1038/s42003-020-0794-7

Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, Carlozzi NE, Bauer PJ, Wallner-Allen K, Fox N, Havlik R, Beaumont JL, Mungas D, Manly JJ, Moy C, Conway K, Edwards E, Nowinski CJ, 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**:567–578. DOI: https://doi.org/10.1017/S1355617714000320, PMID: 24959840

Woolrich MW, Ripley BD, Brady M, Smith SM. 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* **14**:1370–1386. DOI: https://doi.org/10.1006/nimg.2001.0931, PMID: 11707093

Wrigglesworth J, Yaacob N, Ward P, Woods RL, McNeil J, Storey E, Egan G, Murray A, Shah RC, Jamadar SD, Trevaks R, Ward S, Harding IH, Ryan J, ASPREE investigator group. 2022. Brain-predicted age difference is associated with cognitive processing in later-life. *Neurobiology of Aging* **109**:195–203. DOI: https://doi.org/10.1016/j.neurobiolaging.2021.10.007, PMID: 34775210

Zhao X, Zhao XM. 2021. Deep learning of brain magnetic resonance images: a brief review. *Methods* **192**:131–140. DOI: 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* **67**:301–320. DOI: https://doi.org/10.1111/j.1467-9868.2005.00503.x