

Statistical Pitfalls in Brain Age Analyses

Ellyn R. Butler¹, Andrew Chen², Rabie Ramadan³, Trang T. Le⁴, Kosha Ruparel¹, Tyler M. Moore¹, Theodore D. Satterthwaite¹, Fengqing Zhang⁵, Haochang Shou², Ruben C. Gur¹, Thomas E. Nichols^{*6,7}, and Russell T. Shinohara^{+*2}

¹Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

²Penn Statistics in Imaging and Visualization Endeavor, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

³Mathematics Department, Temple University, Philadelphia, PA, 19122, USA

⁴Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, 19104, USA

⁵Department of Psychology, Drexel University, Philadelphia, PA, 19104, USA

⁶Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, OX3 7LF Oxford, United Kingdom

⁷FMRIB, Wellcome Centre for Integrative Neuroimaging

Abstract

Over the past decade, there has been an abundance of research on the difference between age and age predicted using brain features, which is commonly referred to as the “brain age gap”. Researchers have identified that the brain age gap, as a linear transformation of an out-of-sample residual, is dependent on age. As such, any group differences on the brain age gap could simply be due to group differences on age. To mitigate the brain age gap’s dependence on age, it has been proposed that age be regressed out of the brain age gap. If this modified brain age gap (MBAG) is treated as a corrected deviation from age, model accuracy statistics such as R^2 will be artificially inflated. Given the limitations of proposed brain age analyses, further theoretical work is warranted to determine the best way to quantify deviation from normality.

Keywords: BrainAGE, brain age gap, age prediction, residual, deviation, development

Highlights:

- The brain age gap is an out-of-sample residual, and as such varies as a function of age.
- A recently proposed modification of the brain age gap, designed to mitigate the dependence on age, results in inflated model accuracy statistics if used incorrectly.
- Given these limitations, we suggest that new methods should be developed to quantify deviation from normal developmental and aging trajectories.

All code can be found in <https://github.com/PennBBL/brainAgeGapMistake>.

Declarations of interest: none.

* Denotes co-senior authorship

+ Denotes corresponding author

E-mail address: russell.shinohara@penmedicine.upenn.edu

Postal address: Blockley Hall, 2nd Floor, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104-6021

1 Introduction

1 In the past decade, there has been an explosion of research devoted to estimating individuals' ages using
2 features derived from magnetic resonance images (MRIs) of the brain (Franke & Gaser, 2019). From studies
3 using diffusion-weighted features to complex functional connectivity metrics, the literature is extensive (Cole,
4 2020; Erus et al., 2015; Irimia, Torgerson, Goh, & Van Horn, 2015; Li, Satterthwaite, & Fan, 2018; Lin et
5 al., 2016). While age is easily measured through more conventional means, assessing the appearance of the
6 brain with respect to the natural patterns of development and aging provides a framework for dimension
7 reduction; from hundreds of thousands to millions of MRI measurements, these models aim to provide the
8 age of the brain for each subject as a convenient summary measure. The predicted age from these models has
9 been coined "brain age", and the difference between age (sometimes referred to as "chronological age") and
10 brain age is typically referred to as the "brain age gap". Predicted ages are calculated using the following
11 fitted model:

$$\hat{A}_i = \hat{f}(B_{i1}, B_{i2}, \dots, B_{ip}),$$

12 where \hat{A}_i is the predicted age of the i th subject, B_{ij} is the j th brain feature for the i th subject, and $f(\cdot)$ is
13 some function of the brain features.

14 Brain age gap analysis was developed to address two major challenges in neuroscience and medicine:
15 high-dimensionality, and individual risk assessment. Neuroimaging data are high dimensional, with the
16 average T1-weighted scan containing approximately 1,200,000 voxels of brain tissue (Cosgrove, Mazure, &
17 Staley, 2007). Importantly, different parts of the brain follow a variety of trajectories across the lifespan
18 (Coupé, Catheline, Lanuza, & Manjón, 2017; Gennatas et al., 2017; Kennedy et al., 2015). Therefore, in
19 order to better predict age, it is beneficial to use more brain features that complement each other (Varikuti
20 et al., 2018). The main motivation, however, behind brain age gap analyses has been to develop a single
21 number to represent an individual's deviation from some normal trajectory (de Lange & Cole, 2020). This is
22 an admirable goal, since deviating from a normal trajectory may be indicative of or predictive of debilitating
23 disorders (Marquand, Rezek, Buitelaar, & Beckmann, 2016).

24 Researchers often test whether members of a group tend to have their age overestimated compared to
25 a control group, striving to assess whether the disorder is associated with the brain aging prematurely or
26 lagging behind. For instance, Chung et al. (2018) asked if those at clinical high risk for psychosis had a
27 larger brain age gap than healthy controls, and Liem et al. (2017) asked if the brain age gap differed across
28 groups with varying degrees of objective cognitive impairment. Typically, these models are developed using
29 regression or machine learning in one dataset, and are evaluated in a test set. The cross-validation process

30 involves dividing the training set into k folds, estimating the model parameters on $k - 1$ folds, applying the
31 fitted model to the remaining fold, and repeating until every participant in the training set has a predicted
32 age. This procedure helps avoid over-fitting and reporting an inflated model accuracy statistic. Finally, the
33 trained model is applied on a separate test set to predict age of each individual based on their brain features.

34 In this article, we note that the brain age gap, and a recently proposed modified version of it (Beheshti,
35 Nugent, Potvin, & Duchesne, 2019; Chung et al., 2018; Liang, Zhang, & Niu, 2019; Smith, Vidaurre, Alfaro-
36 Almagro, Nichols, & Miller, 2019), are not up to the task of quantifying deviation from a normal trajectory.
37 The brain age gap is a linear transformation of an out-of-sample residual (subsequently referred to as a
38 “prediction error”). As such, it is dependent on the outcome variable (i.e., age) (Le et al., 2018). Therefore,
39 differences in the brain age gap between groups may be due to differences in the brain, or due to differences
40 in the age distributions across groups (Le et al., 2018; Smith et al., 2019). A recently proposed solution to
41 this problem — regressing the brain age gap on age and taking the residuals from this model as a modified
42 brain age gap that is orthogonal to age — creates new problems. In particular, if this new prediction error
43 is treated as a deviation from a subject’s age, which it is not, metrics of model accuracy will be severely
44 inflated.

45 2 Known Limitations of the Brain Age Gap

46 Brain age gap analyses have historically been based on the assumption that the difference between age and
47 predicted age does not vary as a function of age; however, recently several groups have pointed out that this
48 assumption is false (Le et al., 2018; Liang et al., 2019; Smith et al., 2019). Smith et al. (2019) pointed out
49 an extreme case of this error: when age has truly no relationship with brain features, the difference between
50 age and predicted age (“brain age gap”) is a linear function of age, which implies that age explains 100%
51 of the variance in the brain age gap. Smith et al. (2019) note that any subsequent analyses studying the
52 relationship between this gap and other metrics is equivalent to relating a linear transformation of age to
53 other metrics.

54 To flesh out the gravity of this observation, consider an example: If age does not vary as a function of any
55 of the brain parameters, all coefficients, aside from the intercept, will be close to zero with high probability,
56 and the intercept will be close to the mean age of the training sample. Let A_i be the age of the i th subject,
57 $B_{i,j}$ the j th brain feature for the i th subject, ϵ_i random error, and \bar{A} the mean age of the training sample.
58 The brain age model is thus:

$$A_i = \beta_0 + \beta_1 B_{i1} + \beta_2 B_{i2} + \dots + \beta_p B_{ip} + \epsilon_i \quad (1)$$

59 And the fitted values are:

$$\hat{A}_i = \hat{\beta}_0 + \hat{\beta}_1 B_{i1} + \hat{\beta}_2 B_{i2} + \dots + \hat{\beta}_p B_{ip} \approx \bar{A} + 0 \times B_{i1} + 0 \times B_{i2} + \dots + 0 \times B_{ip} = \bar{A}. \quad (2)$$

60 For simplicity, let's assume that the coefficients are estimated to be exactly zero. Suppose the mean age
61 of the training sample is 10 years old. Every person will have an estimated age of 10, so their brain age gap,
62 $\hat{A}_i - A_i$, will be $10 - A_i$. Thus, the brain age gaps are as follows: 15-year-olds have a brain age gap of -5,
63 10-year-olds have a brain age gap of 0, 5-year-olds have a brain age gap of 5, etc. Older participants are
64 estimated as being younger than they are, and younger participants as older. The brain age gap is a linear
65 transformation of a residual (i.e., $\hat{\epsilon}_i = A_i - \hat{A}_i = -(\hat{A}_i - A_i)$), which by definition varies as a function of
66 the outcome variable, in this case age. If the brain features are linearly independent of age, then testing for
67 differences in the brain age gap is equivalent to testing, "Is the mean age of group A different from the mean
68 age of group B?" When testing for differences on the brain age gap in general, the question being asked is
69 similar to "Controlling for the brain features, is the mean age of group A different from the mean age of
70 group B?" Because regression on the residuals of a previous model is not equivalent to multiple regression,
71 this description is not quite correct (Chen, Hribar, & Melessa, 2018; Freckleton, 2002). Thus, interpretation
72 of these residuals is difficult.

73 Even if age varies as a function of the brain parameters, the predicted age for every subject will still be
74 shrunk towards the mean age of the training sample. This is referred to as regression towards the mean, and
75 was first documented by Sir Francis Galton in 1886 (Bland & Altman, 1994). As Liang et al. (2019) noted,
76 this phenomenon is a common feature of many good models. Therefore, older subjects will have negative
77 brain age gap estimates on average simply because they are older, while younger subjects will have positive
78 estimates on average.

79 It is important to note that regression towards the mean is not a failure, but a feature, of regression
80 and related methods. If there is any randomness in a process, observations will tend towards the mean
81 of the outcome variable rather than remain as extreme as they were upon initial sampling (Stigler, 1997).
82 Regression towards the mean is a feature of regression that is actively useful for prediction. Since age is
83 known with certainty, the notion of predicting it makes the construction of a residual awkward. Thus,
84 as we continue to use age prediction as a means to reduce dimensionality, it is important to understand
85 the limitations of using age as an outcome variable and subsequent associated analyses. Recognizing the
86 dependence of the brain age gap on age, researchers have begun to develop methods to mitigate the age-
87 dependence of the brain age gap (Beheshti et al., 2019; Le et al., 2018; Smith et al., 2019). Unfortunately,
88 a misuse of residuals persists, resulting in a systematic overestimation of model accuracy.

89 3 Risks of Using a Modified Brain Age Gap

90 To mitigate the residuals' dependence on age, some researchers apply the following algorithm (Beheshti et
91 al., 2019; Chung et al., 2018; Liang et al., 2019; Smith et al., 2019) (see the appendix for details on Beheshti
92 et al. (2019)'s method). First, a training sample is used to estimate a mapping $f(\cdot)$ from brain features
93 to age. Then, for a left out subject i with brain data $B_{i1}, B_{i2}, \dots, B_{ip}$, the predicted age ("brain age") is
94 estimated as \hat{A}_i :

$$\hat{A}_i = \hat{f}(B_{i1}, B_{i2}, \dots, B_{ip}). \quad (3)$$

95 Then the i th subject's brain age gap (BAG) is

$$\text{BAG}_i = \hat{A}_i - A_i. \quad (4)$$

96 Recognizing the brain age gap's dependence on age, the researcher poses a linear model of the brain age gap
97 on age:

$$\text{BAG}_i = \alpha + \gamma A_i + \delta_i \quad (5)$$

98 where estimated parameters $\hat{\alpha}$ and $\hat{\gamma}$ are found from a regression using training data, and δ_i is random error.
99 Thus, the effect of age is removed, producing the modified brain age gap (MBAG):

$$\text{MBAG}_i = \hat{\delta}_i = \text{BAG}_i - (\hat{\alpha} + \hat{\gamma} A_i), \quad (6)$$

100 which, as the prediction error from model (5), is approximately uncorrelated with age (only exactly uncorre-
101 lated if test data is used to estimate α and γ). Because MBAG has been interpreted as a corrected residual,
102 MBAG is added to (or subtracted from; equivalent in correlation, see Supplement) age. This new variable
103 is then referred to as the corrected predicted age:

$$\hat{A}_i^M = A_i + \text{MBAG}_i = \hat{A}_i + \hat{\alpha} + \hat{\gamma} A_i. \quad (7)$$

104 Because the researcher perceives this predicted age as corrected, they correlate it with age to assess their
105 model's accuracy in predicting age. We will refer to \hat{A}_i^M as the "modified predicted age" and will show below
106 why this age estimation is flawed.

107 MBAG is by no means a more accurate measure of an out-of-sample residual, or prediction error (i.e.,
108 the "brain age gap"). The brain age gap itself is *more* dependent on age the *less* the brain features are

Table 1: Papers reporting inflated model accuracy statistics.

Paper	Before Modification	After Modification
Beheshti et al. (2019)	$\text{Corr}(A, \hat{A})^2 = .38$	$\text{Corr}(A, \hat{A}^M)^2 = .88$
Chung et al. (2018)	$\text{Corr}(A, \hat{A})^2 = .66$	$\text{Corr}(A, \hat{A}^M)^2 = .84$
Liang et al. (2019)	MAE = 1.57	MAE = 1.32
Smith et al. (2019)	$\text{Corr}(A, \hat{A}) = .06$	$\text{Corr}(A, \hat{A}^M) = .99$

Note: MAE = Mean Absolute Error.

109 associated with age. Again, consider the extreme case where age is independent of the brain features. Then,
 110 the brain age gap is *completely determined* by age, as explained in the previous section. If MBAG is treated
 111 as an estimate of the deviation from age, the reported model accuracy (e.g., $\text{Corr}(A_i, \hat{A}_i^M)^2 = R^2$) will
 112 always be inflated relative to the true model accuracy, and often drastically so (see Table 1 for details on
 113 papers that have reported inflated model accuracy statistics). When age has no true dependence on the
 114 brain features, the population covariance between age and predicted age, \hat{A}_i , is zero. But when MBAG is
 115 treated as the deviation from age, $A_i + \text{MBAG}_i$, age and modified predicted age, \hat{A}_i^M , have an approximately
 116 *perfect correlation of 1*.

117 In fact, the inflated correlation can be directly computed as a function of the sample estimates of the
 118 covariance between age and predicted age, the variance of age, and the variance of predicted age (see
 119 Supplement for derivations):

$$\begin{aligned} \text{Corr}(A, \hat{A}^M) &= \frac{-\hat{\gamma}\text{Var}(A) + \text{Cov}(A, \hat{A})}{\sqrt{\text{Var}(A) \times (\text{Var}(\hat{A}) + \hat{\gamma}^2\text{Var}(A) - 2\hat{\gamma}\text{Cov}(A, \hat{A}))}} \\ &= \left(1 + \frac{1}{\left(r_{A\hat{A}} - \hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}}\right)^2} (1 - r_{A\hat{A}}^2)\right)^{-1/2} \end{aligned} \quad (8)$$

120 If $\hat{\alpha}$ and $\hat{\gamma}$ are estimated in the test set, equation (8) can be further simplified:

$$\text{Corr}_{\text{test}}(A, \hat{A}^M) = \left(1 + \frac{\text{Var}(A)}{\text{Var}(\hat{A})} (1 - r_{A\hat{A}}^2)\right)^{-1/2}. \quad (9)$$

121 The equation can be simplified even further if \hat{A} is a linear estimator:

$$\text{Corr}_{\text{test, linear}}(A, \hat{A}^M) = \left(1 + r_{A\hat{A}}^2 (1 - r_{A\hat{A}}^2)\right)^{-1/2}. \quad (10)$$

122 To illustrate the inflated correlation effect and confirm that equation 8 is correct, a series of simulations
 123 were run to compare the transformations that researchers describe performing to the above equation using R

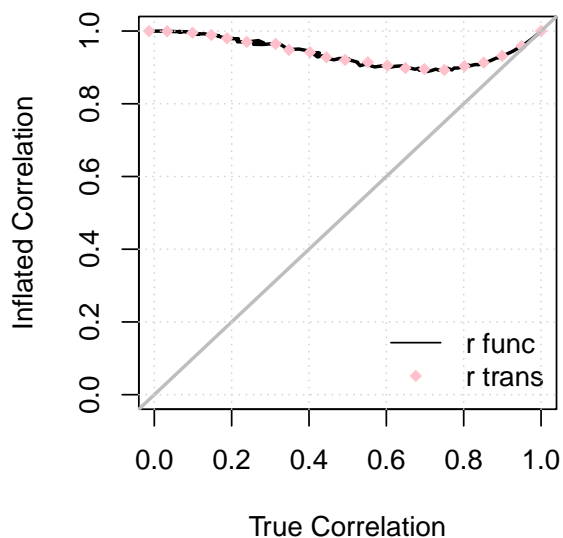


Figure 1: Inflated correlation, $\text{Corr}(A, \hat{A}^M)$, is a function of the true correlation, $\text{Corr}(A, \hat{A})$. The inflated correlation is the correlation between age and the modified predicted age. The true correlation is the correlation between age and predicted age. To illustrate that the series of transformations that researchers perform is equivalent to (8), correlations using both are plotted. r_{func} is using (8), and r_{trans} is using the series of transformations. The identity line is displayed.

124 version 3.6.2 (R Core Team, 2019). Training and testing sets of 10,000 samples were simulated from each of a
125 series of bivariate normal distributions, where the true correlation between age and brain was varied between
126 0 and 1, with the correlation between age and the modified predicted age, \hat{A}_i^M , in the test set being the key
127 outcome measure recorded. All model parameters were estimated in the training set. Since there is only one
128 brain feature, the correlation between age and predicted age is the same as the correlation between age and
129 brain. Results using a single brain feature are detailed in Figure 1. A single brain feature was used so as to
130 have easy control over the correlation between age and predicted age, but note that this result generalizes
131 to any number of brain features. For a set of correlations between 0 and 1, the correlation between age and
132 the modified predicted age, \hat{A}_i^M , was calculated using the theoretical formulation in (8) (black line), and the
133 inflated correlation was obtained using the previously described transformations (pink dots). The identity
134 line is displayed to aid in visualizing that the inflated correlation is larger than the true correlation. The
135 simulations confirmed that the theoretical formulation in (8) is equivalent to the transformations researchers
136 have described. In addition, Figure 1 illustrates that the degree of inflation is much greater for models that
137 have lower values of $\text{Corr}(A, \hat{A})$ than for models that have higher values of $\text{Corr}(A, \hat{A})$.

138 Additional analyses were run using the Philadelphia Neurodevelopmental Cohort (PNC) to illustrate
139 the findings in brain MRI data. Sample details, neuroimaging protocols, and processing can be found in
140 Calkins et al. (2015), Gur et al. (2020), and Satterthwaite et al. (2014). Briefly, participants ages 8-22
141 were recruited through their primary care providers in the Philadelphia area. Subjects were excluded for
142 the purposes of these analyses if their cognitive assessment was conducted more than a year before or after
143 their neuroimaging data was collected, or if their structural image did not pass stringent quality assurance

144 measures. 132 regional volume values were extracted using the Advanced Normalization Tools software
145 package (Tustison et al., 2013; Wang & Yushkevich, 2013).

146 Elastic net models to predict age were built on youths ages 8-22 without a history of mental illness
147 (“typically developing”). Hyperparameters were chosen using repeated five-fold cross validation on the
148 typically developing youth as implemented in the ‘caret’ package, version 6.0-86 (Kuhn, 2012). Then, a
149 linear regression of BAG on age was fit in the typically developing subjects ($N = 317$). Using the fitted
150 values for the parameters from these models, the transformations previously described were applied to youth
151 who met screening criteria for lifetime instance of a mental illness ($N = 862$). This real data example
152 confirmed the theoretical and simulation findings (see Figure 3). Prior to any modification, the correlation
153 between age (A) and predicted age (\hat{A}) was .773. After applying the modifications, the correlation became
154 .884. There were no differences between the typically developing youth and youth with a history of mental
155 illness on age ($t = -1.05, p = 0.29$), the brain age gap ($t = 0.72, p = 0.47$), or MBAG ($t = 0.09, p = 0.390$).
156 Age and performance on the complex cognition tasks were highly associated ($r = 0.54, p < .0001$). After
157 regressing the brain features out of age and multiplying by negative one – or constructing the brain age gap
158 – this association weakened ($r = -0.30, p < .0001$). MBAG and performance on the complex cognition tasks
159 were not associated ($r = -.01, p = 0.71$). These results indicate that the association between cognition
160 and the brain age gap are driven by the association between age and cognitive performance. Prior work
161 highlighting group differences and correlations between brain age metrics and other variables should be
162 examined in light of these results.

163 4 Conclusion

164 We have shown that predicted age estimates (“brain age”) based on a regression adjustment of the brain age
165 gap result in a correlation between a modified predicted age and age never falling much below 0.9 regardless
166 of the original predicted age and age correlation. Further, the interpretability of MBAG itself is limited by
167 the fact that it is a prediction error from a regression to remove the effects of age from a residual obtained
168 through a regression to predict age. By virtue of these limitations, we suggest that the brain age gap and
169 the modified version may not provide useful information about precocity or delay in brain development. In
170 light of this, we suggest that methods be developed specifically to answer questions about similarity between
171 brains of different age groups and diseased states.

172 Many other transformations have been developed to mitigate the downstream effects of BAG’s dependence
173 on age (de Lange & Cole, 2020). Some are not susceptible to the inflated correlation issue described in this
174 work. Methods include scaling the predicted age by the slope and intercept from the regression of predicted

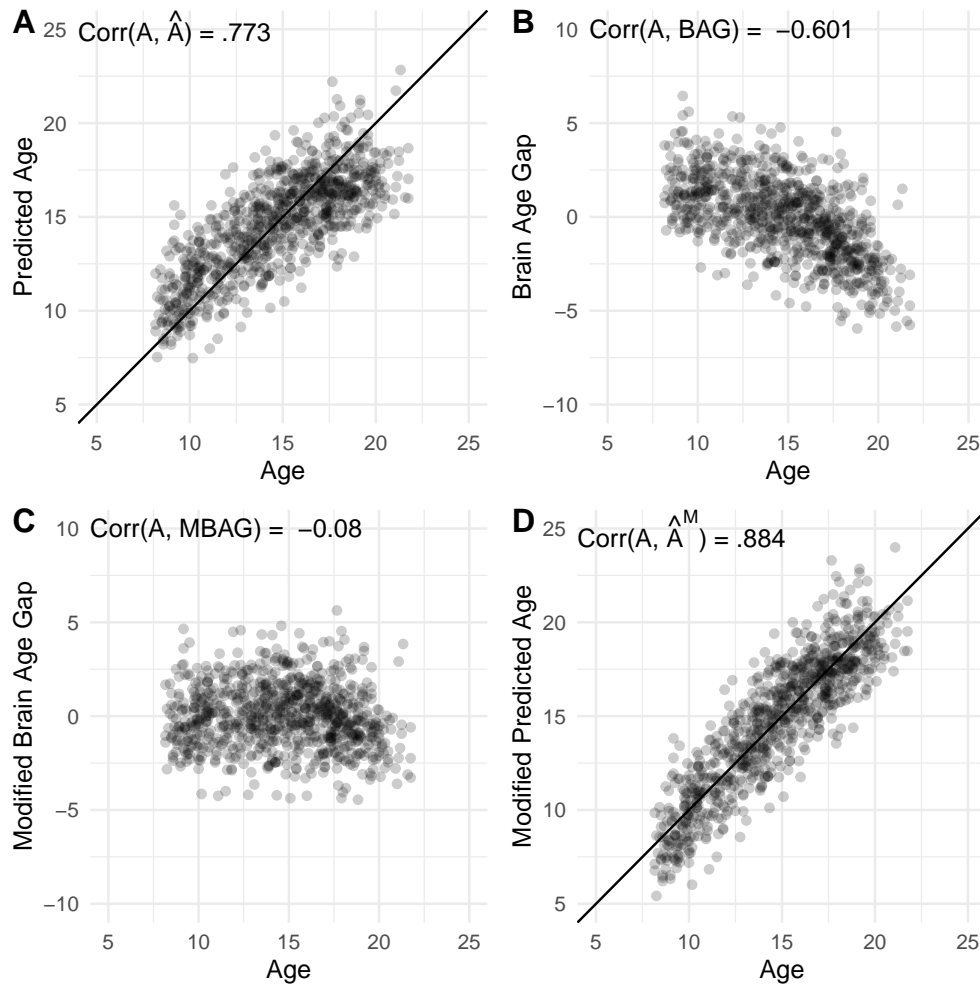


Figure 2: The inflated correlation finding was replicated in the Philadelphia Neurodevelopmental Cohort. Plotted are values for age, predicted age, brain age gap, modified brain age gap and modified predicted age in the subset of participants who met screening criteria for an instance of mental illness in their lifetime. The identity line is displayed in panels A and D.

175 age on age (see (5) in de Lange and Cole (2020)), and including age as a covariate when testing for group
176 differences in BAG (Le et al., 2018). The former results in a new BAG estimate that is uncorrelated with
177 age, and the latter ensures that any group differences found on BAG will be linearly independent of age.
178 Note that, if all models had been built on the test set, controlling for age when testing for group differences
179 on BAG is the two-step regression equivalent of including age as a covariate in a multiple regression with
180 brain features predicting age. The real question then becomes: to what extent do these methods quantify
181 advanced or delayed brain development? This question warrants further theoretical investigation.

182 Future research should also determine appropriate analytic methods to answer whether the brains of
183 patients with disorders are more similar to older healthy controls' than age-matched healthy controls' brains,
184 and to evaluating the extent to which analyses of residuals as deviations from some trajectory exist in the
185 literature. Thus far, we are aware of a similar trend of predicting age using genetic features in attempt to
186 document differences in precocious and delayed genetic development (Sumner, Colich, Uddin, Armstrong,
187 & McLaughlin, 2019; Wolf et al., 2018). In the meantime, while previous studies have suggested that the
188 brain age gap be used as biomarker in clinical trials (Cole et al., 2018), our findings suggest that further
189 methodological work is warranted.

190 **5 Acknowledgements**

191 E.R.B. conceptualized the work, wrote derivations, and wrote the manuscript. A.C., T.E.N., and R.T.S.
192 wrote derivations, and provided feedback on the framing of the work. R.R., T.T.L., F.Z. and H.S. provided
193 statistical insights and edited the manuscript. T.D.S. oversaw the processing of the neuroimaging data and
194 edited the manuscript. K.R., T.M.M., and R.C.G. provided inspiration for, and assisted in the framing of
195 the work. We thank Dr. Michael Stein, members of the Brain Behavior Laboratory and attendees of Data
196 Club for their questions and feedback.

197 Funding: This work was supported by the National Institute of Mental Health grant numbers MH107235,
198 MH117014, MH119219, MH123550, and MH112847; and National Institute of Neurological Disorders and
199 Stroke grant numbers NS060910 and NS112274.

References

- Beheshti, I., Nugent, S., Potvin, O., & Duchesne, S. (2019). Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. *NeuroImage: Clinical*, *24*, 102063.
- Bland, J. M., & Altman, D. G. (1994). Regression towards the mean. *BMJ (Clinical research ed.)*, *308*(6942), 1499.
- Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., . . . others (2015). The philadelphia neurodevelopmental cohort: constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry*, *56*(12), 1356–1369.
- Chen, W., Hribar, P., & Melessa, S. (2018). Incorrect Inferences When Using Residuals as Dependent Variables. *Journal of Accounting Research*, *56*(3), 751–796. doi: 10.1111/1475-679X.12195
- Chung, Y., Addington, J., Bearden, C. E., Cadenhead, K., Cornblatt, B., Mathalon, D. H., . . . Cannon, T. D. (2018). Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. *JAMA Psychiatry*, *75*(9), 960–968. doi: 10.1001/jamapsychiatry.2018.1543
- Cole, J. H. (2020). Multi-modality neuroimaging brain-age in uk biobank: relationship to biomedical, lifestyle and cognitive factors. *Neurobiology of Aging*.
- Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., . . . Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, *23*(5), 1385–1392. Retrieved from <http://dx.doi.org/10.1038/mp.2017.62> doi: 10.1038/mp.2017.62
- Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving Knowledge of Sex Differences in Brain Structure. *Biological psychiatry*, *62*(8), 847–855. doi: 10.1016/j.biopsych.2007.03.001.Evolving
- Coupé, P., Catheline, G., Lanuza, E., & Manjón, J. V. (2017). Towards a unified analysis of brain maturation and aging across the entire lifespan: A MRI analysis. *Human Brain Mapping*, *38*(11), 5501–5518. doi: 10.1002/hbm.23743
- de Lange, A.-M. G., & Cole, J. H. (2020). Commentary: Correction procedures in brain-age prediction. *NeuroImage: Clinical*, *26*.
- Erus, G., Battapady, H., Satterthwaite, T. D., Hakonarson, H., Gur, R. E., Davatzikos, C., & Gur, R. C. (2015). Imaging patterns of brain development and their relationship to cognition. *Cerebral Cortex*, *25*(6), 1676–1684. doi: 10.1093/cercor/bht425
- 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*(JUL). doi: 10.3389/fneur.2019.00789
- Freckleton, R. P. (2002). On the misuse of residuals in ecology: Regression of residuals vs. multiple regression.

- 232 *Journal of Animal Ecology*, 71(3), 542–545. doi: 10.1046/j.1365-2656.2002.00618.x
- 233 Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., . . . Gur, R. C.
234 (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical
235 thickness from childhood to young adulthood. *Journal of Neuroscience*, 37(20), 5065–5073. doi:
236 10.1523/JNEUROSCI.3550-16.2017
- 237 Gur, R. C., Butler, E. R., Moore, T. M., Rosen, A. F., Ruparel, K., Satterthwaite, T. D., . . . Gur,
238 R. E. (2020). Structural and functional brain parameters related to cognitive performance across
239 development: Replication and extension of the parieto-frontal integration theory in a single sam-
240 ple. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2020/07/21/659193> doi:
241 10.1101/659193
- 242 Irimia, A., Torgerson, C. M., Goh, S.-Y. M., & Van Horn, J. D. (2015). Statistical estimation of physiological
243 brain age as a descriptor of senescence rate during adulthood. *Brain imaging and behavior*, 9(4), 678–
244 689.
- 245 Kennedy, K. M., Rodrigue, K. M., Bischof, G. N., Hebrank, A. C., Reuter-Lorenz, P. A., & Park,
246 D. C. (2015). Age trajectories of functional activation under conditions of low and high process-
247 ing demands: An adult lifespan fMRI study of the aging brain. *NeuroImage*, 104, 21–34. doi:
248 10.1016/j.neuroimage.2014.09.056
- 249 Kuhn, M. (2012). The caret package. *R Foundation for Statistical Computing, Vienna, Austria*. URL
250 [https://cran.r-project.org/package= caret](https://cran.r-project.org/package=caret).
- 251 Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., & Paulus, M. P. (2018). A
252 Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers*
253 *in Aging Neuroscience*, 10(October), 1–11. doi: 10.3389/fnagi.2018.00317
- 254 Li, H., Satterthwaite, T. D., & Fan, Y. (2018). Brain age prediction based on resting-state functional
255 connectivity patterns using convolutional neural networks. In *2018 IEEE 15th International Symposium*
256 *on Biomedical Imaging (ISBI 2018)* (pp. 101–104).
- 257 Liang, H., Zhang, F., & Niu, X. (2019, aug). Investigating systematic bias in brain age estimation with
258 application to post-traumatic stress disorders. *Human Brain Mapping*, 40(11), 3143–3152. doi:
259 10.1002/hbm.24588
- 260 Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J. M., . . . Margulies,
261 D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment.
262 *NeuroImage*, 148(July 2016), 179–188. doi: 10.1016/j.neuroimage.2016.11.005
- 263 Lin, L., Jin, C., Fu, Z., Zhang, B., Bin, G., & Wu, S. (2016). Predicting healthy older adult’s brain age based
264 on structural connectivity networks using artificial neural networks. *Computer methods and programs*

265 *in biomedicine*, 125, 8–17.

266 Marquand, A. F., Rezek, I., Buitelaar, J., & Beckmann, C. F. (2016). Understanding Heterogeneity in
267 Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biological Psychiatry*, 80(7),
268 552–561. doi: 10.1016/j.biopsych.2015.12.023

269 R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual].
270 Vienna, Austria. Retrieved from <https://www.R-project.org/>

271 Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., ... others
272 (2014). Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage*, 86, 544–553.

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

275 Stigler, S. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical
276 Research*, 6, 103–114.

277 Sumner, J. A., Colich, N. L., Uddin, M., Armstrong, D., & McLaughlin, K. A. (2019). Early experiences of
278 threat, but not deprivation, are associated with accelerated biological aging in children and adolescents.
279 *Biological psychiatry*, 85(3), 268–278.

280 Tustison, N. J., Avants, B. B., Cook, P. A., Song, G., Das, S., van Strien, N., ... Gee, J. C. (2013). The ants
281 cortical thickness processing pipeline. In *Medical imaging 2013: Biomedical applications in molecular,
282 structural, and functional imaging* (Vol. 8672, p. 86720K).

283 Varikuti, D. P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K. R., ... Eickhoff, S. B.
284 (2018). Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*,
285 173(March), 394–410. doi: 10.1016/j.neuroimage.2018.03.007

286 Wang, H., & Yushkevich, P. (2013). Multi-atlas segmentation with joint label fusion and corrective learn-
287 ing—an open source implementation. *Frontiers in neuroinformatics*, 7, 27.

288 Wolf, E. J., Maniates, H., Nugent, N., Maihofer, A. X., Armstrong, D., Ratanatharathorn, A., ... others
289 (2018). Traumatic stress and accelerated dna methylation age: a meta-analysis. *Psychoneuroen-
290 docrinology*, 92, 123–134.

291 **Beheshti et al. (2019) correlation**

292 Beheshti et al. (2019) suggest subtracting $\hat{\alpha} + \hat{\gamma}A_i$ from \hat{A}_i , and calling this new value the corrected
 293 predicted age:

$$\begin{aligned} \text{Corr}(A, \hat{A} - (\hat{\alpha} + \hat{\gamma}A)) &= \text{Corr}(A, \hat{A} - \hat{\gamma}A) \\ &= \frac{\text{Cov}(A, \hat{A} - \hat{\gamma}A)}{\sqrt{\text{Var}(A)\text{Var}(\hat{A} - \hat{\gamma}A)}} \\ &= \frac{-\hat{\gamma}\text{Var}(A) + \text{Cov}(A, \hat{A})}{\sqrt{\text{Var}(A)(\text{Var}(\hat{A}) + \hat{\gamma}^2\text{Var}(A) - 2\hat{\gamma}\text{Cov}(A, \hat{A}))}} \\ &= \text{Corr}(A, \hat{A}^M). \end{aligned}$$

294

295 Therefore, their method is equivalent to Eqn. 8.

296

297 **Adding and subtracting MBAG from age results in the same**
 298 **inflated correlation with age**

Modified predicted age has been calculated in the literature as either $\text{MBAG}_i = A_i - \text{BAG}_i$ or $\text{MBAG}_i = A_i + \text{BAG}_i$. In both cases, the main results from the paper applies since the correlation between age and the modified predicted age using either formula is the same. We have that

$$\begin{aligned} \text{Corr}(A, A - \text{MBAG}) &= \frac{\text{Var}(A) - \text{Cov}(A, \text{MBAG})}{\sqrt{\text{Var}(A)(\text{Var}(A) + \text{Var}(\text{MBAG}) - 2\text{Cov}(A, \text{MBAG}))}} \\ &= \frac{\text{Var}(A)}{\sqrt{\text{Var}(A)^2 + \text{Var}(A)\text{Var}(\text{MBAG})}} \end{aligned}$$

$$\begin{aligned} \text{Corr}(A, A + \text{MBAG}) &= \frac{\text{Var}(A) + \text{Cov}(A, \text{MBAG})}{\sqrt{\text{Var}(A)(\text{Var}(A) + \text{Var}(\text{MBAG}) + 2\text{Cov}(A, \text{MBAG}))}} \\ &= \frac{\text{Var}(A)}{\sqrt{\text{Var}(A)^2 + \text{Var}(A)\text{Var}(\text{MBAG})}} \\ &= \text{Corr}(A, A - \text{MBAG}) \end{aligned}$$

299 which follows from the fact that MBAG_i is a residual from regression of BAG_i on A_i and thus MBAG_i is
 300 orthogonal to A_i or equivalently, $\text{Cov}(A, \text{MBAG}) = 0$. Note that this result is only approximate when the
 301 regression of BAG on age is done in the training set.

302

Derivation of Equation (8)

303

$$\begin{aligned}
 \text{Corr}(A, \hat{A}^M) &= \text{Corr}(A, A + \text{MBAG}) \\
 &= \text{Corr}(A, A + \text{BAG} - (\hat{\alpha} + \hat{\gamma}A)) \\
 &= \text{Corr}(A, \hat{A} - \hat{\alpha} - \hat{\gamma}A) \\
 &= \frac{\text{Cov}(A, \hat{A} - \hat{\alpha} - \hat{\gamma}A)}{\sqrt{\text{Var}(A)\text{Var}(\hat{A} - \hat{\alpha} - \hat{\gamma}A)}}, \text{ apply the definition of correlation} \\
 &= \frac{-\hat{\gamma}\text{Var}(A) + \text{Cov}(A, \hat{A})}{\sqrt{\text{Var}(A)\left(\text{Var}(\hat{A}) + \hat{\gamma}^2\text{Var}(A) - 2\hat{\gamma}\text{Cov}(A, \hat{A})\right)}} \\
 &= \frac{-\hat{\gamma}\frac{\text{Var}(A)}{\sqrt{\text{Var}(A)\text{Var}(\hat{A})}} + \frac{\text{Cov}(A, \hat{A})}{\sqrt{\text{Var}(A)\text{Var}(\hat{A})}}}{\sqrt{\frac{1}{\text{Var}(A)\text{Var}(\hat{A})}\left[\text{Var}(A)\left(\text{Var}(\hat{A}) + \hat{\gamma}^2\text{Var}(A) - 2\hat{\gamma}\text{Cov}(A, \hat{A})\right)\right]}} \\
 &= \frac{-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + \text{Corr}(A, \hat{A})}{\sqrt{1 + \hat{\gamma}^2\frac{\text{Var}(A)}{\text{Var}(\hat{A})} - 2\hat{\gamma}\frac{\text{Cov}(A, \hat{A})}{\text{Var}(\hat{A})}} \\
 &= \frac{-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + r_{A\hat{A}}}{\sqrt{1 + \left(\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2 - 2\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}}r_{A\hat{A}}}} \\
 &= \frac{-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + r_{A\hat{A}}}{\sqrt{1 + \left(\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2 - 2\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}}r_{A\hat{A}} + r_{A\hat{A}}^2 - r_{A\hat{A}}^2}} \\
 &= \frac{-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + r_{A\hat{A}}}{\sqrt{1 - r_{A\hat{A}}^2 + \left(r_{A\hat{A}} - \hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2}}, \text{ factor the quadratic} \\
 &= \left(\left(-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + r_{A\hat{A}}\right)^{-2}\right)^{-1/2} \left(1 - r_{A\hat{A}}^2 + \left(r_{A\hat{A}} - \hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2\right)^{-1/2} \\
 &= \left(\frac{1 - r_{A\hat{A}}^2 + \left(r_{A\hat{A}} - \hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2}{\left(-\hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}} + r_{A\hat{A}}\right)^2}\right)^{-1/2} \\
 &= \left(1 + \frac{1}{\left(r_{A\hat{A}} - \hat{\gamma}\sqrt{\frac{\text{Var}(A)}{\text{Var}(\hat{A})}\right)^2}(1 - r_{A\hat{A}}^2)\right)^{-1/2}
 \end{aligned}$$

304

305 Derivations of Equations (9) and (10)

306

The following derivation involves the algebraic manipulation of the sample estimates and not expectations. Assuming that the linear regression of BAG on age has been estimated with the testing data, then

$$\begin{aligned} \text{MBAG} &= \text{BAG} - H_A \text{BAG} \\ &= \hat{A} - A - H_A(\hat{A} - A) \\ &= \hat{A} - A - H_A \hat{A} + H_A A \\ &= \hat{A} - H_A \hat{A} - (A - H_A A) \\ &= \hat{A} - H_A \hat{A} \\ &= R_A \hat{A} \end{aligned}$$

307 where $H_A = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ is the hat matrix for the regression on age, $\mathbf{A} = [\mathbf{1} \ A]$, and $R_A = I - H_A$ is the
308 corresponding residual forming matrix.

309 We first note that

$$\begin{aligned} \text{Cov}(A, \hat{A}^M) &= \text{Cov}(A, A + \text{MBAG}) \\ &= \text{Cov}(A, A + R_A \hat{A}) \\ &= \text{Var}(A) + \text{Cov}(A, R_A \hat{A}) \\ &= \text{Var}(A) \end{aligned}$$

310 where the last equality holds due to the orthogonality of A and R_A .

Then, consider:

$$\begin{aligned} \text{Var}(\hat{A}^M) &= \text{Var}(A + \text{MBAG}) \\ &= \text{Var}(A + R_A \hat{A}) \\ &= \text{Var}(\hat{A}) \left(\frac{\text{Var}(A)}{\text{Var}(\hat{A})} + \frac{\text{Var}(R_A \hat{A})}{\text{Var}(\hat{A})} \right) \\ &= \text{Var}(\hat{A}) \left(\frac{\text{Var}(A)}{\text{Var}(\hat{A})} + (1 - r_{A\hat{A}}^2) \right). \end{aligned}$$

311 Then, Eqn. (9) is found as

$$\begin{aligned}
 \text{Corr}(A, \hat{A}^M) &= \frac{\text{Var}(A)}{\sqrt{\text{Var}(A)\text{Var}(\hat{A}^M)}} \\
 &= \frac{\text{Var}(A)}{\sqrt{\text{Var}(A)\text{Var}(\hat{A}) \left[\frac{\text{Var}(A)}{\text{Var}(\hat{A})} + (1 - r_{A\hat{A}}^2) \right]}} \\
 &= \left[\frac{\text{Var}(\hat{A})}{\text{Var}(A)} \left(\frac{\text{Var}(A)}{\text{Var}(\hat{A})} + (1 - r_{A\hat{A}}^2) \right) \right]^{-1/2} \\
 &= \left(1 + \frac{\text{Var}(\hat{A})}{\text{Var}(A)} (1 - r_{A\hat{A}}^2) \right)^{-1/2}.
 \end{aligned}$$

For insight on the $\text{Var}(\hat{A})/\text{Var}(A)$ term, note that shrinkage will generally mean this term is less than one. Moreover, if \hat{A} were found with a linear regression on the testing data, i.e. $\hat{A} = X(X^T X)^{-1} X^T A$ where X are brain features, then this ratio is exactly the squared correlation,

$$\frac{\text{Var}(\hat{A})}{\text{Var}(A)} = r_{A\hat{A}}^2,$$

312 producing Eqn. (10).

313

314 In this setting, when both brain age and MBAG are determined from testing data using linear regression,
 315 the correlation of A and \hat{A}^M can never fall below $\frac{1}{\sqrt{1+0.5^2*(1-0.5^2)}} \approx 0.9177$. Of course, in practice, held-out
 316 training data is used to learn the brain-age relationship, so a regression prediction would instead have the
 317 form $\hat{A} = X(X_{\text{in}}^T X_{\text{in}})^{-1} X_{\text{in}}^T A_{\text{in}}$, where X_{in} and A_{in} are held-in training data, but $\text{Var}(\hat{A})/\text{Var}(A) \approx r_{A\hat{A}}^2$ still
 318 provides a useful starting point for exploring the parameters in the expression for $\text{Corr}(A, \hat{A}^M)$.

319

320 Finally, note that the equality of $\text{Var}(\hat{A})/\text{Var}(A)$ and $r_{A\hat{A}}^2$ holds not just for linear regression, but any lin-
 321 ear estimator. Specifically, if there exists an idempotent H_X ($H_X H_X = I, H_X^T = H_X$) such that $\mathbf{1}^T H_X = \mathbf{1}$
 322 and $\hat{A} = H_X A$, then

323

$$\begin{aligned}\frac{\text{Var}(\hat{A})}{\text{Var}(A)} &= \frac{\text{Var}(H_X A)}{\text{Var}(A)} \\ &= \frac{(H_X A)^T / N - (\mathbf{1}^T H_X A / N)^2}{\text{Var}(A)} \\ &= \frac{A^T H_X A / N - (\mathbf{1}^T A / N)(\mathbf{1}^T H_X A / N)}{\text{Var}(A)} \\ &= \frac{\text{Cov}(A, H_X A)}{\text{Var}(A)} \\ &= \frac{\text{Cov}(A, \hat{A})}{\text{Var}(A)} \\ &= \sqrt{\frac{\text{Var}(\hat{A})}{\text{Var}(A)}} r_{A\hat{A}}.\end{aligned}$$

And thus

$$\frac{\text{Var}(\hat{A})}{\text{Var}(A)} = r_{A\hat{A}}^2.$$