Statistical Pitfalls in Brain Age Analyses

Ellyn R. Butler[1], Andrew Chen[2], Rabie Ramadan[3], Trang T. Le[4], Kosha Ruparel[1], Tyler M. Moore[1], Theodore D. Satterthwaite[1], Fengqing Zhang[5], Haochang Shou[2], Ruben C. Gur[1], Thomas E. Nichols[*6,7], and Russell T. Shinohara[+*2]

[1]Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
[2]Penn Statistics in Imaging and Visualization Endeavor, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
[3]Mathematics Department, Temple University, Philadelphia, PA, 19122, USA
[4]Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, 19104, USA
[5]Department of Psychology, Drexel University, Philadelphia, PA, 19104, USA
[6]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, OX3 7LF Oxford, United Kingdom
[7]FMRIB, 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@pennmedicine.upenn.edu
Postal address: Blockley Hall, 2nd Floor, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104-6021

1 Introduction

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

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

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

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

Researchers often test whether members of a group tend to have their age overestimated compared to a control group, striving to assess whether the disorder is associated with the brain aging prematurely or lagging behind. For instance, Chung et al. (2018) asked if those at clinical high risk for psychosis had a larger brain age gap than healthy controls, and Liem et al. (2017) asked if the brain age gap differed across groups with varying degrees of objective cognitive impairment. Typically, these models are developed using regression or machine learning in one dataset, and are evaluated in a test set. The cross-validation process involves dividing the training set into k folds, estimating the model parameters on $k-1$ folds, applying the fitted model to the remaining fold, and repeating until every participant in the training set has a predicted age. This procedure helps avoid over-fitting and reporting an inflated model accuracy statistic. Finally, the trained model is applied on a separate test set to predict age of each individual based on their brain features.

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

2 Known Limitations of the Brain Age Gap

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

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

$$A_i = \beta_0 + \beta_1 B_{i1} + \beta_2 B_{i2} + \cdots + \beta_p B_{ip} + \epsilon_i \tag{1}$$

And the fitted values are:

$$\hat{A}_i = \hat{\beta}_0 + \hat{\beta}_1 B_{i1} + \hat{\beta}_2 B_{i2} + \cdots + \hat{\beta}_p B_{ip} \approx \overline{A} + 0 \times B_{i1} + 0 \times B_{i2} + \cdots + 0 \times B_{ip} = \overline{A}. \tag{2}$$

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

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

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

3 Risks of Using a Modified Brain Age Gap

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

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

Then the ith subject's brain age gap (BAG) is

$$\mathrm{BAG}_i = \hat{A}_i - A_i. \tag{4}$$

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

$$\mathrm{BAG}_i = \alpha + \gamma A_i + \delta_i \tag{5}$$

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

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

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

$$\hat{A}_i^M = A_i + \mathrm{MBAG}_i = \hat{A} + \hat{\alpha} + \hat{\gamma} A_i. \tag{7}$$

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

MBAG is by no means a more accurate measure of an out-of-sample residual, or prediction error (i.e., 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.

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

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

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

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

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

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

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

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

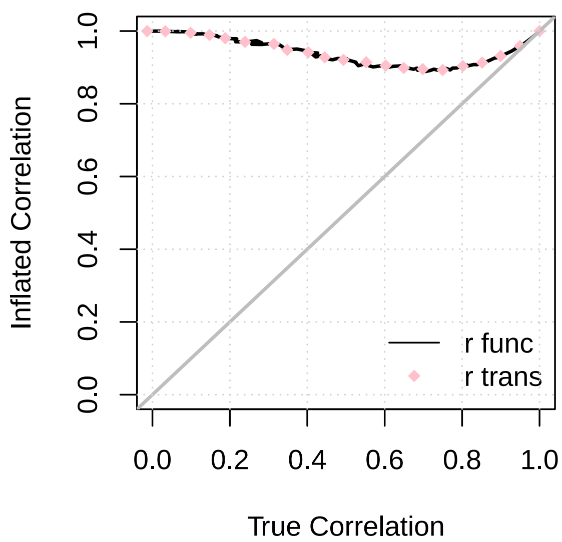

Figure 1: Inflated correlation, $\mathrm{Corr}(A, \hat{A}^M)$, is a function of the true correlation, $\mathrm{Corr}(A, \hat{A})$. The inflated correlation is the correlation between age and the modified predicted age. The true correlation is the correlation between age and predicted age. To illustrate 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.

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

Additional analyses were run using the Philadelphia Neurodevelopmental Cohort (PNC) to illustrate the findings in brain MRI data. Sample details, neuroimaging protocols, and processing can be found in Calkins et al. (2015), Gur et al. (2020), and Satterthwaite et al. (2014). Briefly, participants ages 8-22 were recruited through their primary care providers in the Philadelphia area. Subjects were excluded for the purposes of these analyses if their cognitive assessment was conducted more than a year before or after their neuroimaging data was collected, or if their structural image did not pass stringent quality assurance measures. 132 regional volume values were extracted using the Advanced Normalization Tools software package (Tustison et al., 2013; Wang & Yushkevich, 2013).

Elastic net models to predict age were built on youths ages 8-22 without a history of mental illness ("typically developing"). Hyperparameters were chosen using repeated five-fold cross validation on the typically developing youth as implemented in the 'caret' package, version 6.0-86 (Kuhn, 2012). Then, a linear regression of BAG on age was fit in the typically developing subjects (N = 317). Using the fitted values for the parameters from these models, the transformations previously described were applied to youth who met screening criteria for lifetime instance of a mental illness (N = 862). This real data example confirmed the theoretical and simulation findings (see Figure 3). Prior to any modification, the correlation between age (A) and predicted age ($\hat{A}$) was .773. After applying the modifications, the correlation became .884. There were no differences between the typically developing youth and youth with a history of mental illness on age ($t = -1.05, p = 0.29$), the brain age gap ($t = 0.72, p = 0.47$), or MBAG ($t = 0.09, p = 0390$). Age and performance on the complex cognition tasks were highly associated ($r = 0.54$, $p < .0001$). After regressing the brain features out of age and multiplying by negative one – or constructing the brain age gap – this association weakened ($r = -0.30$, $p < .0001$). MBAG and performance on the complex cognition tasks were not associated ($r = -.01$, $p = 0.71$). These results indicate that the association between cognition and the brain age gap are driven by the association between age and cognitive performance. Prior work highlighting group differences and correlations between brain age metrics and other variables should be examined in light of these results.

4 Conclusion

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

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

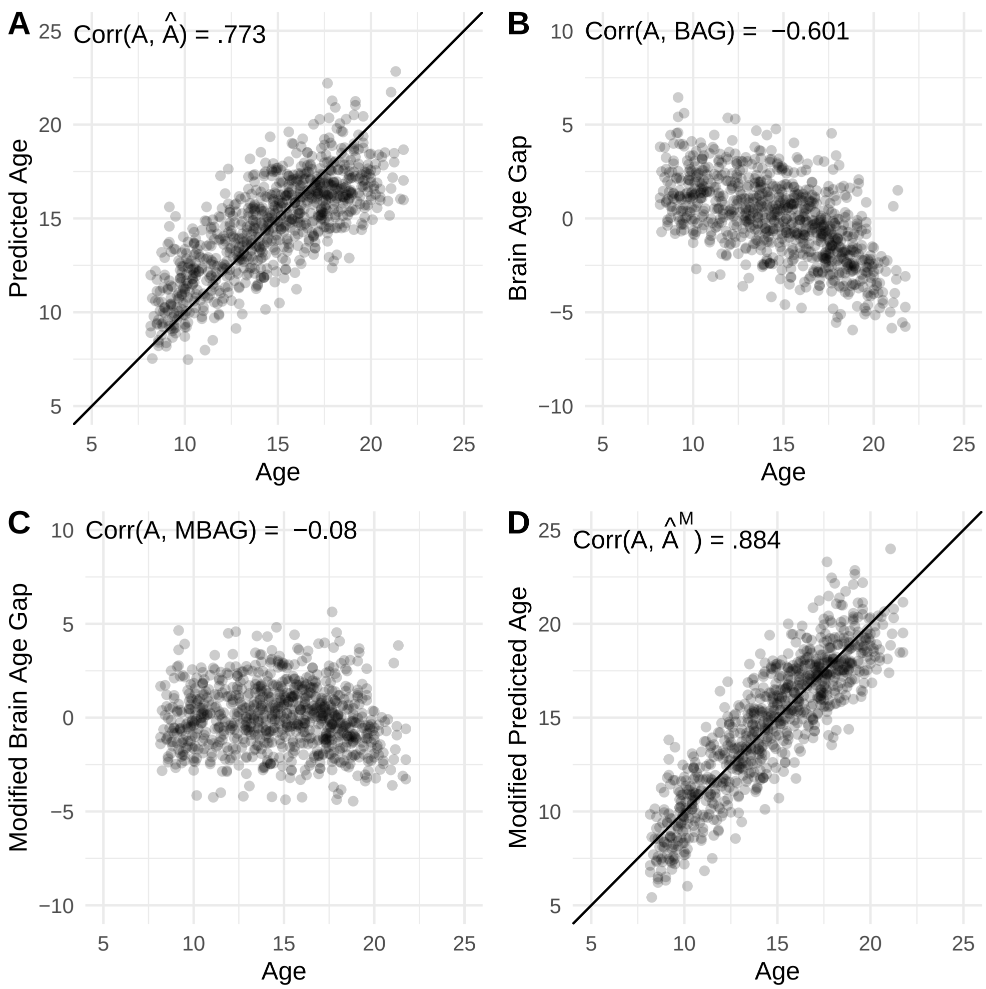

Figure 2: The inflated correlation finding was replicated in the Philadelphia Neurodevelopmental Cohort. Plotted are values for age, predicted age, brain age gap, modified brain age gap and modified predicted age in the subset of participants who met screening criteria for an instance of mental illness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

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

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

5 Acknowledgements

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

Funding: This work was supported by the National Institute of Mental Health grant numbers MH107235, MH117014, MH119219, MH123550, and MH112847; and National Institute of Neurological Disorders and 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). Evoling 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.

Journal of Animal Ecology, *71*(3), 542–545. doi: 10.1046/j.1365-2656.2002.00618.x

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., … Gur, R. C. (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *Journal of Neuroscience*, *37*(20), 5065–5073. doi: 10.1523/JNEUROSCI.3550-16.2017

Gur, R. C., Butler, E. R., Moore, T. M., Rosen, A. F., Ruparel, K., Satterthwaite, T. D., … Gur, R. E. (2020). Structural and functional brain parameters related to cognitive performance across development: Replication and extension of the parieto-frontal integration theory in a single sample. *bioRxiv*. Retrieved from https://www.biorxiv.org/content/early/2020/07/21/659193 doi: 10.1101/659193

Irimia, A., Torgerson, C. M., Goh, S.-Y. M., & Van Horn, J. D. (2015). Statistical estimation of physiological brain age as a descriptor of senescence rate during adulthood. *Brain imaging and behavior*, *9*(4), 678–689.

Kennedy, K. M., Rodrigue, K. M., Bischof, G. N., Hebrank, A. C., Reuter-Lorenz, P. A., & Park, D. C. (2015). Age trajectories of functional activation under conditions of low and high processing demands: An adult lifespan fMRI study of the aging brain. *NeuroImage*, *104*, 21–34. doi: 10.1016/j.neuroimage.2014.09.056

Kuhn, M. (2012). The caret package. *R Foundation for Statistical Computing, Vienna, Austria. URL https://cran. r-project. org/package= caret*.

Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., & Paulus, M. P. (2018). A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in Aging Neuroscience*, *10*(October), 1–11. doi: 10.3389/fnagi.2018.00317

Li, H., Satterthwaite, T. D., & Fan, Y. (2018). Brain age prediction based on resting-state functional connectivity patterns using convolutional neural networks. In *2018 ieee 15th international symposium on biomedical imaging (isbi 2018)* (pp. 101–104).

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

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J. M., … Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, *148*(July 2016), 179–188. doi: 10.1016/j.neuroimage.2016.11.005

Lin, L., Jin, C., Fu, Z., Zhang, B., Bin, G., & Wu, S. (2016). Predicting healthy older adult's brain age based on structural connectivity networks using artificial neural networks. *Computer methods and programs* *in biomedicine*, *125*, 8–17.

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

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

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

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

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

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

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

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

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

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

Beheshti et al. (2019) correlation

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

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

Therefore, their method is equivalent to Eqn. 8.

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

Moditified predicted age has been calculated in the literature as either $\mathrm{MBAG}_i = A_i - \mathrm{BAG}_i$ or $\mathrm{MBAG}_i = A_i + \mathrm{BAG}_i$. 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

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

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

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

Derivation of Equation (8)

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

Derivations of Equations (9) and (10)

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

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

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 corresponding residual forming matrix.

We first note that

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

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

Then, consider:

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

Then, Eqn. (9) is found as

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

For insight on the $\mathrm{Var}(\hat{A})/\mathrm{Var}(A)$ 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. $\hat{A} = X(X^TX)^{-1}X^TA$ where X are brain features, then this ratio is exactly the squared correlation,

$$\frac{\mathrm{Var}(\hat{A})}{\mathrm{Var}(A)} = r^2_{A\hat{A}},$$

producing Eqn. (10).

In this setting, when both brain age and MBAG are determined from testing data using linear regression, 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 training data is used to learn the brain-age relationship, so a regression prediction would instead have the form $\hat{A} = X(X_{\mathrm{in}}^T X_{\mathrm{in}})^{-1} X_{\mathrm{in}}^T A_{\mathrm{in}}$, where X_{in} and A_{in} are held-in training data, but $\mathrm{Var}(\hat{A})/\mathrm{Var}(A) \approx r^2_{A\hat{A}}$ still provides a useful starting point for exploring the parameters in the expression for $\mathrm{Corr}(A, \hat{A}^M)$.

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

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

And thus

$$
\frac{\mathrm{Var}(\hat{A})}{\mathrm{Var}(A)} = r^2_{A\hat{A}}.
$$