Associations between smoking and accelerated brain ageing

Zeqiang Linli a,b, Jianfeng Feng c,d, Wei Zhao a,b,*, Shuixia Guo a,b,*

a MOE-LCSM, School of Mathematics and Statistics, Hunan Normal University, Changsha 410006, PR China
b Key Laboratory of Applied Statistics and Data Science, Hunan Normal University, College of Human Province, Changsha 410006, PR China
c Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
d Centre for Computational Systems Biology, Fudan University, Shanghai 200433, PR China

* Corresponding author at: School of Mathematics and Statistics, Hunan Normal University, Changsha 410006, PR China.
E-mail addresses: 18182105361@163.com (W. Zhao), guoshuixia75@163.com (S. Guo).

https://doi.org/10.1016/j.pnpbp.2021.110471
Received 19 August 2021; Received in revised form 27 October 2021; Accepted 31 October 2021
Available online 3 November 2021
0278-5846/© 2021 Elsevier Inc. All rights reserved.

ARTICLE INFO

Keywords:
Ageing
Brain age
Cognition
Gray matter
Smoking

ABSTRACT

Smoking accelerates the ageing of multiple organs. However, few studies have quantified the association between smoking, especially smoking cessation, and brain ageing. Using structural magnetic resonance imaging data from the UK Biobank (n = 33,293), a brain age predictor was trained using a machine learning technique in the non-smoker group (n = 14,667) and then tested in the smoker group (n = 18,626) to determine the relationships between BrainAge Gap (predicted age – true age) and smoking parameters. Further, we examined whether smoking was associated with poorer cognition and whether this relationship was mediated by brain age. The predictor achieved an appreciable performance in training data (r = 0.712, mean-absolute-error [MAE] = 4.220) and test data (r = 0.725, MAE = 4.160). On average, smokers showed a larger BrainAge Gap (+0.304 years, Cohens’d = 0.083) than controls, more explicitly, the extents vary depending on their smoking characteristic that active regular smokers had the largest BrainAge Gap (+1.190 years, Cohens’d = 0.321), and light smokers had a moderate BrainAge Gap (+0.478, Cohens’d = 0.129). The increased smoking amount was associated with a larger BrainAge Gap (β = 0.035, p = 1.72 × 10^-20) while a longer duration of quitting smoking in ex-smokers was associated with a smaller BrainAge Gap (β = -0.015, p = 2.14 × 10^-05). Furthermore, smoking was associated with poorer cognition, and this relationship was partially mediated by BrainAge Gap. The study provides insight into the association between smoking, brain ageing, and cognition, which provide more publicly acceptable propaganda against smoking.

1. Introduction

Over the decades, smoking has become one of the biggest threats to world health. There are about 1.3 billion smokers worldwide in 2019, accounting for about 16% of the world population, and about 8 million people die from smoking every year (WHO, 2020). Smoking accelerates the process of organ ageing and leads to multiple diseases, such as circulatory and respiratory diseases (Jha et al., 2013; Walters et al., 2014; Wu et al., 2019). Furthermore, smoking may lead to multiple neurodegenerative and neurobiological abnormalities such as cognitive decline and dementia (Debette et al., 2011; Durazzo et al., 2014; Frisoni et al., 2010; Vnuková et al., 2017), suggesting that smoking is associated with accelerated brain ageing. Although many studies have investigated the group-level association between smoking and brain structure, which located many but heterogeneous smoking-related potential differences (Ebbejani et al., 2019; Gray et al., 2020; Yang et al., 2020), few studies have quantified the association between smoking with brain ageing in detail. An individual-specific measure conceptualizing brain ageing of every smoker intuitively and straightforwardly would be helpful to reflect the overall brain health, identify people at risk of age-associated health problems, develop personalized therapy for smoking cessation, and more importantly, to provide more publicly acceptable propaganda against smoking.

Recently, a “brain age estimate” paradigm based on neuroimaging, especially on structural magnetic resonance imaging (MRI), has been successfully used to examine advanced or delayed brain ageing in healthy and clinical populations (Cole and Franke, 2017; Franke and Gaser, 2019). The mathematical rationale for this paradigm has been proven (Niu et al., 2020; Smith et al., 2019). The typical brain age estimation framework is as follows: (i) Imaging pre-processing, in which the raw neuroimaging is preprocessed, and features are identified for use in the following analysis steps. (ii) Model training, in which the pre-processing imaging features (i.e., independent variables) in a training dataset of healthy people, along with their chronological age (i.e.,
dependent variable), are then entered into a machine learning regression algorithm. iii) Model evaluation, in which the model is evaluated using the resample strategy, commonly using a cross-validation procedure in which some proportion of the individuals in the training set are left out of the initial training stage, which is used to determine the accuracy of the model. The metrics frequently used in regression include the mean absolute error (MAE), root mean square error (RMSE), and the Pearson coefficient $r$ between chronological and estimated brain age. iv) Model generalization, in which the final brain age prediction model is applied to previously unseen test subjects (e.g., clinical group) to estimate the subject-specific brain ages using their imaging features. Subtracting chronological age from estimated brain age quantifies the acceleration or deceleration of individual brain ageing, which is called the BrainAge Gap (Cole et al., 2015; Koutsouleris et al., 2014; Peng et al., 2021). Intuitively, a positive gap represents an older-appearing brain, while a negative gap represents a younger-appearing brain. This index has been widely applied in various neurological and psychiatric disorders, including schizophrenia, major depressive disorder, traumatic brain injury, Alzheimer’s disease, and mild cognitive impairment (Cole et al., 2015; Gaser et al., 2013; Han et al., 2020; Koutsouleris et al., 2014; Liem et al., 2017), indicating a higher brain age for participants in the clinical group than their chronological age.

Smoking, just like neuropsychiatric disorders, increases the BrainAge Gap (Bittner et al., 2021; Cole, 2020; Ning et al., 2020; Smith et al., 2020). However, it is critical to note that most of these published studies focused on the differences between current smokers or ever-smokers and nonsmokers. Indeed, smoking behaviors influenced by a variety of individual factors are so complex that it is overly simplistic to divide the population into smokers and non-smokers. Analyzing smokers in a broader way based on measures of their smoking could provide a deeper understanding of the association of smoking and brain age, which of course needs a larger study population. Although smoking quantified in pack-years is associated with accelerated brain ageing (Bittner et al., 2021), no studies have quantified the association between smoking cessation and brain ageing. Understanding the association between smoking amount and cessation and brain ageing has important public health and treatment implications.

This study aimed to investigate the relationship between three smoking parameters (smoking status, amount of smoking, and smoking cessation) and brain ageing and cognitive functions in a middle-aged and older population using data from the UK Biobank, the largest neuroimaging databases. We hypothesized that i) a relatively high accuracy would be achieved for the brain age predictor using an advanced machine-learning technique, ii) smoking parameters would be significantly associated with the BrainAge Gap, and iii) smoking would be associated with poorer cognition, and this relationship could be mediated by brain age.

2. Materials and methods

2.1. Participants

This study included 38,562 middle-aged and old adults (44 to 81 years old) with structural MRI. The UK Biobank received ethical approval from the research ethics committee (REC reference 11/NW/0382). Written informed consent was obtained from each subject. Data access permission was granted under UKB application 19,542 (PI Jianfeng Feng).

With the purpose of our study, participants were excluded if they had (1) reported neuropsychological disorder such as bipolar disorder, depression, and mania; (2) reporting missing or unclear smoking data (for example, smoking cigarettes on most or all days reported missing or unclear started smoking ages); (3) reporting missing key demographic variables (e.g., age); (4) poor quality of sMRI. Following exclusions, there were 33,293 subjects with structural MRI data included in the following analysis.

2.2. Imaging data collection and preprocessing

The UK Biobank used a standard Siemens Skyra 32-channel 3 T scanner (Siemens Medical Solutions, Germany) for all magnetic-resonance brain imaging, with $1 \times 1 \times 1$ resolution and a view field of $208 \times 256 \times 256$. The details of the image acquisition are provided at the UK Biobank website in the form of a protocol (http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=2367).

All UK Biobank structural MRI data were preprocessed in the CAT12 toolbox with default settings, including: (1) the T1-weighted images were segmented into GM, white matter (WM), and non-brain voxels (cerebrospinal fluid, skull) using the “new-segment” routine; (2) population templates (GM, WM) were generated from each of the dataset separately using the DARTEL algorithm; (3) the gray-matter images were aligned to a nonlinear deformation field and normalized to MNI space; (4) the normalized images were then smoothed with an 8 mm full-width at half-maximum Gaussian kernel with the resulting voxel size 1.5mm$^3$.Spatially normalized, smoothed, and Jacobian scaled gray-matter images were obtained for each subject. The estimated total intraventricular volume (TIV) was calculated as the summation of the gray matter, white matter, and cerebrospinal fluid volumes in the native space. The automated anatomical labeling 3 (AAL3) atlas, which partitioned the brain into 166 regions of interest, was employed to obtain region-wise gray matter volume (Rolls et al., 2020). Detailed labels of anatomical regions in the AAL3 atlas were listed in Table S1.

2.3. Research variables

2.3.1. Smoking parameters

We divided the 33,293 subjects into the smoker ($n = 18,626$) and never-smoking control ($n = 14,667$) groups according to whether they have ever smoked. Furthermore, smokers were split into 6 groups based on their smoking characteristics (Table S2). Specifically, those who currently smoke on most or all days were classified as “Current smoker” ($n = 1254$); those who previously smoked on most or all days but occasionally still smoked were classified as “Ex-smoker1” ($n = 240$); those who previously smoked on most or all days and had stopped smoking at the time of the study were classified as “Ex-smoker2” ($n = 6749$); those who had smoked occasionally in the past and now and had smoked a total of at least 100 times were classified as “Light smoker1” ($n = 418$); those who used to smoke occasionally and smoked a total of at least 100 times but quit smoking currently were classified as “Light smoker2” ($n = 3751$); and those who had extremely lightly smoked and failed to meet the standards of light smokers (i.e., a total of at least 100 times in their lifetime) were classified as “Light smokerX” ($n = 6214$).

The other two key smoking parameters calculated were: i) pack-years was calculated as cigarettes per day divided by 20 and then times the number of years smoked and was only available for current and ex-smokers; ii) quitting duration, which was calculated as age at the time of data collection minus the age when the participant stopped smoking on most days and this was only available for ex-smokers.

2.3.2. Cognitive function

The UK Biobank also contains a series of cognitive measures (https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100026). The present study included 5 cognitive measures with continuous test scores, namely reaction time (Filed ID: 20023), fluid intelligence (Filed ID: 20016), numeric memory (Filed ID: 4282), pairs matching (Filed ID: 399), and symbol digit substitution (Filed ID: 20159).

Reaction time gives a crude measure of the raw processing and reaction speed of a participant that the larger the value, the slower the reaction speed (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20023).

Fluid intelligence is a simple unweighted sum of the number of correct answers given to the 13 fluid intelligence questions (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20016).
Numeric memory assesses the numeric short-term memory of a participant. The values of this cognitive function used in the present study were the longest number correctly recalled during the numeric memory test that the larger the value, the better the memory (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4282).

Pairs matching test asked participants to memorize the position of as many matching pairs of cards as possible. The values of this cognitive function used here were the number of incorrect matches in the round that the larger the value, the worse the memory. (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=399).

Symbol digit substitution is a neuropsychological test. The values of this cognitive test used here were the numbers of symbols correctly matched to digits by the participant that the larger the value, the better the cognition. (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=23324).

2.3.3. Mental health

Given that anxiety, unhappiness, and mania are the common mental health symptoms for smokers, these measures were included in this study. These variables were calculated according to the data categories 136, 140, and 147 in the UK Biobank website, and were converted into binary variables based on their medians (e.g., higher anxiety vs. lower anxiety). Detailed information on these variables can be found in the Supplementary Materials.

2.3.4. Other covariables

Variables as follows theoretically likely to relate to brain structure and cognitive function were included as covariates in the following analyses: age, sex, handedness, ethnicity, body-mass index (BMI), alcohol consumption, and TIV.

2.4. Analysis overview

Fig. 1 outlines the analytical pipeline used in this study. After pre-processing, we residualized all 166 regional GMVs for sex, ethnicity, handedness, BMI, scanning site, alcohol consumption, and TIV using linear regression models. To train a brain-age prediction model, subjects were divided into smokers and never-smoking controls. The control...
group trained the predictors using a nested 5-fold cross-validation (CV) framework. The final predictors with optimal parameters were used to predict the brain age of the smoker group. Next, to correct for the systematic bias in brain age prediction leading to generally overestimated age predictions at low age and underestimated predictions at high age, a correcting model was used across all controls, and the model coefficients were then used to remove the brain age bias among smokers. Finally, corrected brain age was used in subsequent analyses, including comparative analysis, association analysis, and mediation analysis.

2.5. GMV based brain age prediction framework

In this study, extreme gradient boosting (XGBoost) predictors, a popular technique in recent brain age studies (Kaufmann et al., 2019; de Lange et al., 2019; de Lange et al., 2020), was used to build a brain-age estimating model across all controls, with chronological age as the dependent variable and 166 regional GMV as independent variables, and was implemented using the R package xgboost (https://github.com/mlc/xgboost). XGBoost is an implementation of gradient-boosted trees designed for speed and performance; the final model is based on a collection of individual models. Compared to regular gradient boosting, which uses the loss function of the base model as a proxy for minimizing error, XGBoost computes second-order derivatives to provide information about the direction of gradients and how to obtain the minimum loss function. It also includes advanced regularization to reduce overfitting (Chen and Guestrin, 2016).

We used a nested five-fold CV framework, with the outer five-fold CV loop to estimate the generalizability of the model and the inner five-fold CV loop to determine the optimal parameter set for the XGBoost regressor using the grid search method over the search space. In the outer five-fold CV, all controls were randomly divided into five subsets stratified by age to preserve the same age distribution. Within each loop of the outer five-fold CV, we applied an inner five-fold CV, that is, the training set for each loop of the outer five-fold CV was further partitioned into five subsets. Four subsets were selected to train the model under a given parameter set, and the remaining subset was used to test the model. In this study, the tuning parameter ranges were set to maximum depth \( M = \{2, 3, 4, 5, 6, 7, 8, 9, 10\} \), number of estimators \( N = \{50, 100, 150, 200, 250\} \), and learning rate \( \eta = \{0.01, 0.05, 0.1, 0.2, 0.5\} \), and the default values were used for all other parameters. For each set of \( (M, N, \eta) \), the mean squared error was measured for each inner five-fold CV loop, and a mean value for all the five-fold inner loops was then obtained to indicate the inner prediction performance. The parameter set with the highest inner prediction accuracy among the inner five-fold CVs was chosen as the optimal parameter. Finally, all samples of the inner five-fold CV were trained with the best parameters, and the testing subjects of the outer five-fold CV were predicted. The training and testing procedures of the outer five-fold CV were repeated five times, with each subset used once as the testing set.

Therefore, a full nested five-fold CV loop finally produced five XGBoost predictors with optimal parameters, and all training data (controls) had a predicted age (interpreted as BrainAge). The BrainAge Gap can be computed by subtracting the chronological age from the estimated brain age (i.e., BrainAge – True age). The BrainAge of smokers was predicted using five XGBoost predictors and then averaged across the five predicted values as the final BrainAge of smokers, and the BrainAge Gap was also calculated for every smoker.

To investigate the prediction accuracy, Pearson correlation coefficients \( r \), MAE, and RMSE calculated between the predicted age and chronological age were calculated for the control and smoker groups, respectively.

2.6. Brain age correction

Due to problems such as regression dilution and non-Gaussian age distribution (Le et al., 2018; Smith et al., 2019), brain-age estimation involves a frequently observed bias: BrainAge is overestimated in younger subjects and underestimated in older subjects, while BrainAge for participants with an age closer to the mean age (of the training dataset) is predicted more accurately. To avoid spurious correlations between BrainAge Gap and age-related variables in the following analysis, a bias adjustment procedure was performed. A common practice is to apply a regression model to remove the effect of chronological age on the estimated BrainAge (Beheshti et al., 2019; Liang et al., 2019; Smith et al., 2019). We used a correcting model across all controls using the following formula:

\[
\text{BrainAge} = \beta_0 + \beta_1 \times \text{age} + \beta_2 \times \text{age}^2 + \epsilon
\]

where the residuals of the regression eq. (1) represent the difference between brain age and chronological age after controlling for confounding factors, including the linear and quadratic effects of chronological age. We referred to this as the corrected BrainAge Gap. The corrected BrainAge can be calculated using the following formulas:

Corrected BrainAge Gap = BrainAge Gap \cdot \left( \frac{\hat{\beta}_0 + \hat{\beta}_1 \times \text{age} + \hat{\beta}_2 \times \text{age}^2}{\beta_0 + \beta_1 \times \text{age} + \beta_2 \times \text{age}^2} \right)

Corrected BrainAge = \text{Chronological Age} + \text{Corrected BrainAge Gap}

The regression coefficients of this correction model were then used for bias removal in the smoker group.

2.7. Sensitivity analysis

To ensure that the BrainAge or BrainAge Gap we used above is robust, we conducted several sensitivity analyses.

(1) As described in the Introduction section, the general pipeline of brain age prediction is to train a model from a large sample of healthy people and then apply it to those with neuropsychiatric or physical health conditions. However, some studies trained predictors using all subjects (de Lange et al., 2019; de Lange et al., 2020). To examine whether this different pipeline affects the prediction performance and subsequent statistical results, we retrained the XGBoost predictors using the whole dataset (smokers + never-smoking controls).

(2) To prove the generalizability of the brain-age model we established across the different countries and even different ethnic groups, we applied the brain-age prediction models trained on a never-smoking control group from the UK Biobank to an independent test set \( (N = 492) \) from China, the Southwest University Adult Lifespan Dataset (SALD) (Wei et al., 2018), which is crucial for broad application in a clinical context, and investigate the potential power of the brain age models as a diagnostic and prediction tool at a single-subject level.

2.8. Variable importance in XGBoost predictor

A benefit of using gradient boosting is that after the boosted trees are constructed, it is relatively straightforward to retrieve the importance of each variable, providing a sense of which features are powering the predictions. Formulas were developed by estimating the relative influence of variables according to a previous study estimate the relative influence of variables (Friedman, 2001). In brief, the measures of importance are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Elith et al., 2008; Friedman and Meulman, 2003). In this study, variable importance was calculated for each of the 166 brain regions using the R function varimp and averaged across five XGBoost predictors and then scaled to the interval 0–1, with higher values indicating a stronger influence on the response.
status was significantly associated with the BrainAge Gap by correcting 2.9.1. Comparative analysis using linear regression, with the cognitive measure as the dependent bias unless stated otherwise. Three statistical analyses were performed. mediation model was implemented using the R package mediation (ht t://CRAN.R-project.org/package=mediation)(Baron and Kenny, 1986; Wager et al., 2008). Mediation analysis tests that investigate whether the association between the independent variable X (smoking variable) and the dependent variable Y (cognitive function) can be explained by the mediation variable M (BrainAge Gap). Specifically, estimates were calculated for the total effect of the smoking variable on cognition (X → Y), the effect of the smoking variable on the BrainAge Gap (X → M), and the effect of the BrainAge Gap on cognitions adjusting for the smoking variable (X → M → Y). The significance of the mediation, that is, whether the relationship had been significantly reduced with the inclusion of the mediator, was estimated using the bias-corrected bootstrap approach (with 1000 random samplings). The percentage of the mediation effect that could be explained by the mediator (indirect effect) was measured using the formula: 100 × (total effect – direct effect)/total effect. Confounding variables, as in the association analysis, were regressed out in the mediation model. Like any other regression analysis, mediation analysis does not imply causal relationships.

3. Results

3.1. Demographic characteristics

Of 33,293 participants, the mean age at enrolment was 63.73 (SD = 7.53) years. 15,651 (47.0%) were male and 1038 (3.1%) were non-white people. The cohort included 1254 (3.77%) current smokers, 6749 (20.27%) ex-smokers1, 240 ex-smokers2 (0.72%), 3751 light smokers1 (11.27%), 418 light smokers2 (1.26%), 6214 light smokersX (18.66%) and 14,667 controls (44.05%). Smokers showed the worst mental health condition; compared to controls, a significantly larger percentage of smokers had relatively more anxiety (49.0% vs. 49.5%–64.1%), lower well-being (49.0% vs. 57.5%–73.3%), and higher mania status (23.8% vs. 25.8%–41.9%), with the magnitude related to smoking frequency and smoking cessation. Detailed characteristics of these groups are provided in Table 1.

Table 1: Demographic variables of different groups.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Overall (n = 33,293)</th>
<th>Current Smoker (n = 1254)</th>
<th>Ex-smoker1 (n = 240)</th>
<th>Ex-smoker2 (n = 6749)</th>
<th>Light Smoker1 (n = 3751)</th>
<th>Light Smoker2 (n = 6214)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean (SD))</td>
<td>63.73 (7.53)</td>
<td>61.84 (7.29)</td>
<td>62.59 (7.37)</td>
<td>65.64 (7.21)</td>
<td>61.72 (7.64)</td>
<td>64.36 (7.48)</td>
</tr>
<tr>
<td>Male (%)</td>
<td>15,651 (47.0%)</td>
<td>644 (51.4%)</td>
<td>130 (54.2)</td>
<td>3557 (52.72)</td>
<td>259 (62.0)</td>
<td>1918 (51.1)</td>
</tr>
<tr>
<td>Handedness: non-right (%)</td>
<td>3671 (11.0)</td>
<td>146 (11.6)</td>
<td>28 (11.7)</td>
<td>783 (11.6)</td>
<td>39 (9.3)</td>
<td>440 (11.7)</td>
</tr>
<tr>
<td>Ethnic: non-white (%)</td>
<td>1038 (3.1)</td>
<td>46 (3.7)</td>
<td>11 (4.6)</td>
<td>135 (2.0)</td>
<td>16 (3.8)</td>
<td>107 (2.9)</td>
</tr>
<tr>
<td>Pairs matching</td>
<td>0.38 (0.92)</td>
<td>0.39 (0.90)</td>
<td>0.35 (0.86)</td>
<td>0.38 (0.92)</td>
<td>0.40 (0.85)</td>
<td>0.40 (0.99)</td>
</tr>
<tr>
<td>Symbol digit substitution</td>
<td>20.35 (4.98)</td>
<td>19.84 (5.18)</td>
<td>20.09 (4.78)</td>
<td>19.70 (4.90)</td>
<td>20.46 (4.81)</td>
<td>20.00 (4.80)</td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher anxiety (%)</td>
<td>11,576 (34.5)</td>
<td>518 (41.4)</td>
<td>110 (63.6)</td>
<td>2360 (50.0)</td>
<td>143 (51.8)</td>
<td>1263 (49.5)</td>
</tr>
<tr>
<td>Lower wellbeing (%)</td>
<td>13,290 (39.8)</td>
<td>602 (47.3)</td>
<td>115 (68.0)</td>
<td>2881 (62.5)</td>
<td>168 (63.2)</td>
<td>1435 (57.5)</td>
</tr>
<tr>
<td>Higher mania (%)</td>
<td>6031 (26.4)</td>
<td>343 (26.8)</td>
<td>72 (41.9)</td>
<td>1328 (28.3)</td>
<td>90 (32.7)</td>
<td>706 (27.8)</td>
</tr>
</tbody>
</table>

Group comparison p-values were calculated based on the variable categories, that is, chi-square tests were used for categorical variables (with continuity correction) and analyses of variance were used for continuous variables. The meaning of NA is not applicable. There are different available sample sizes for each mental health score or cognitive measure in analysis.

2.9. Statistical analysis

The BrainAge Gap used in the statistical analysis was adjusted for bias unless stated otherwise. Three statistical analyses were performed.

2.9.1. Comparative analysis

Two-sample two-tailed t-tests were used to test whether smoking status was significantly associated with the BrainAge Gap by correcting for the confounding effects of age, sex, handedness, ethnicity, BMI, alcohol consumption, and TIV. We also compared five cognition measures (reaction time, fluid intelligence, numeric memory, pair matching, and symbol digit substitution) of smokers to those of controls using a linear model with covariates including age, sex, handedness, ethnicity, BMI, alcohol consumption, and TIV. In addition, we investigated whether worse mental health conditions were associated with a larger BrainAge Gap using two-sample t-tests.

2.9.2. Association analysis

We used a linear model to investigate the association between two continuous smoking parameters (pack-years and duration of quitting) and the BrainAge Gap, with smoking parameters as independent variables and the BrainAge Gap as the dependent variable, covariates including age, sex, handedness, ethnicity, BMI, alcohol consumption, and TIV. Similarly, the association between these two smoking parameters and the five cognitive measures was examined. We also investigated the associations between BrainAge Gap and cognitive functions using linear regression, with the cognitive measure as the dependent variable and the BrainAge Gap as the independent variable, with adjustments for potential confounders.

2.9.3. Mediation analysis

To verify whether the relationship between smoking and cognition was mediated by the BrainAge Gap, a standard three-variable path mediation model was implemented using the R package mediation (ht p://CRAN.R-project.org/package=mediation)(Baron and Kenny, 1986;
3.2. Brain age prediction performance in the control group

The brain age prediction models that were run on the control group using XGBoost implemented in a nested five-fold CV framework showed a prediction accuracy of $r = 0.712$, CI = 0.703–0.719 (Fig. 2A), RMSE = 5.280, and MAE = 4.220. The BrainAge Gap for every control (Fig. 2B) showed a negative association with chronological age, as expected ($r = -0.707$). After bias adjusting using linear regression with the formula BrainAge = 51.68 + 0.151 $\times$ age + 0.005 $\times$ age$^2$, the corrected BrainAge of controls was correlated more with chronological age ($r = 0.896$, CI = 0.893–0.899) and the corrected BrainAge Gap was orthogonal to the chronological age ($r = 0$; Fig. 2D).

3.3. Contributing brain regions in brain age prediction model

The brain imaging feature importance obtained from the XGBoost model is shown in Fig. 3. The most important ROIs in brain age estimation were located in regions such as the temporal lobe (the fusiform gyrus, middle temporal gyrus, and hippocampus), lingual gyrus, and thalamus (the lateral geniculate) (Table S3). Moreover, the more important the region was in prediction, the stronger it was in the age-GMV association (Spearman’s $r = -0.206$, $p < 0.001$; Fig. S1).

3.4. Generalization of brain age prediction model to the smoker groups

The predictors trained from the control group were used to predict the BrainAge of the smoker group. The correlation between the predicted and chronological age in the smoker group was $r = 0.725$, CI = 0.718–0.732, RMSE = 5.179, and MAE = 4.160 (Fig. S2A). Similarly, the BrainAge Gap was also correlated with the chronological age ($r = -0.706$, Fig. S2B). Using the correlation coefficients of the model from the control group to remove the BrainAge bias of smokers, the corrected BrainAge of the smoker group was more relevant to chronological age than before the correction with $r = 0.900$, CI = 0.896–0.903 (Fig. S2C), RMSE = 3.664, and MAE = 2.943, and the corrected BrainAge Gap was nearly orthogonal to chronological age ($r = 0.014$, $p = 0.05$; Fig. S2D). Further, the correlation between the corrected BrainAge and chronological age in the six smoking groups is shown in Fig. 4, with $r = 0.9$ and $p < 0.001$ for all groups.

3.5. Comparison of BrainAge gap between different groups

As illustrated in Table 2, smokers had a bigger BrainAge Gap than controls, with a mean difference (MD) of 0.304 years, $p < 0.001$, CI = 0.223–0.384, and Cohen’s $d$ = 0.083. To further quantify the difference in the BrainAge Gap of smokers and controls, each smoking subgroup was compared with the control group. As shown in Fig. 5A and Table 2, current smokers had the largest BrainAge Gap, with MD = 1.190 years, $p < 0.001$, and Cohen’s $d$ = 0.233, Ex-smoker1 group (MD = 0.859 years, $p < 0.001$, and Cohen’s $d$ = 0.323), Ex-smoker2 group (MD = 0.552 years, $p < 0.001$, and Cohen’s $d$ = 0.149), and Light smoker1 (MD = 0.478 years, $p = 0.009$, and Cohen’s $d$ = 0.129). There were no significant differences in the BrainAge Gap between the Light smoker2 group and Light smokerX group and the Control group.

In addition, participants with higher anxiety ($p = 0.001$), lower well-being ($p = 0.036$), and higher mania ($p = 0.004$) status had a larger BrainAge Gap (Table S4).

3.6. Association of Smoking Variables with BrainAge

There was a positive association between BrainAge Gap and smoking amount (pack-years) in ever-smokers (Current, Ex-smoker1, and Ex-smoker2) with reference to controls (pack-years = 0) ($\beta$ = 0.027, $r = 0.088$, $p = 2.73 \times 10^{-46}$; Table 3 and Fig. SB). The coefficient was greater if only current smokers were considered ($\beta = 0.035$, $r = 0.073$, $p = 1.69 \times 10^{-20}$). The opposite trend was found in the association between BrainAge Gap and duration of quitting smoking in former smokers (Ex-smoker1 and Ex-smoker2) with reference to current smokers (quitting duration = 0), with $\beta = -0.025$, $r = -0.094$, and $p = 1.77 \times 10^{-17}$ (Table 3 and Fig. SC), while after controlling the effect of pack-years, the effect size of BrainAge Gap duration was weakened ($\beta = -0.015$, $r = -0.042$, and $p = 2.12 \times 10^{-05}$; Fig. 5D) shows that the BrainAge Gap was generally larger than that of control if smokers with more pack-years and that this association was stronger for those whose duration of quitting was shorter or zero.

3.7. Association of Smoking Variables with cognitive function

Among the five cognitive functions, reaction time and symbol digit substitution were significantly associated with smoking variables (Table 4). Compared with the control group, the “Current smoker” group had significantly longer reaction times ($\beta = 9.464$, $q = 7.54 \times 10^{-03}$), and reaction time positively correlated with the number of pack-years ($\beta = 0.350$, $q = 3.01 \times 10^{-03}$) and negatively correlated with quitting duration ($\beta = -0.258$, $q = 7.79 \times 10^{-03}$). Regarding the symbol digit substitution scores, the “Current smoker” and “Ex-smoker2” groups showed smaller scores than the control group ($\beta = -1.155$, $q = 2.71 \times 10^{-08}$ and $\beta = -0.313$, $q = 9.86 \times 10^{-03}$, respectively), and significant associations with the number of pack-years and quitting duration were found ($\beta = -0.039$, $q = 1.24 \times 10^{-07}$ and $\beta = 0.017$, $q = 1.07 \times 10^{-02}$, respectively).
3.8. Association of BrainAge gap with cognitive function

To investigate whether the BrainAge Gap was associated with cognitive function, we explored the associations between the BrainAge Gap and five cognitive measures. Table 5 shows that a larger BrainAge gap correlated with a longer reaction time ($\beta = 0.408, q = 2.23 \times 10^{-3}$), lower fluid intelligence scores ($\beta = 0.018, q = 6.39 \times 10^{-3}$), more incorrect pair matching number ($\beta = 0.004, q = 2.16 \times 10^{-2}$), and lower symbol digit substitution scores ($\beta = 0.091, q = 4.30 \times 10^{-21}$). These correlations were greater when only smokers were considered (Table 5).

3.9. BrainAge gap mediates the association between smoking variables and cognitive function

Given the association between BrainAge Gap and both symbol digit substitution scores and smoking parameters, we investigated whether the BrainAge Gap differences associated with smoking parameters are related to the impairments in symbol digit substitution ability associated with smoking. The results indicated that the BrainAge Gap significantly mediated the relationship between the smoking variables and symbol digit substitution (Fig. 6), that is, BrainAge Gap partially mediated the association between smoking status ($\beta = -0.0689, p < 0.001, CI = -0.1032\text{--}0.0376$, and proportion of mediation = 5.8%), pack-years ($\beta = -0.0025, p < 0.001, CI = -0.0037\text{--}0.0014$, and proportion of mediation = 6.4%), and quitting duration ($\beta = 0.0012, p = 0.002, CI = 0.0005\text{--}0.0023$, and proportion of mediation = 7.1%) and symbol digit substitution scores.

3.10. Results of sensitive analysis

The detailed results of the sensitivity analysis are described in the Supplementary Materials. In brief, using all participants (controls + smokers) to train brain age estimators, the predicted age was highly correlated with that using only the control group (similar to the main study), regardless of correcting the brain age bias ($r = 0.967, p < 0.001$ and $r = 0.986, p < 0.001$, respectively; Fig. S3). Furthermore, the association between the BrainAge Gap with smoking parameters was similar to that observed in the main experiment (Fig. S4).

Since the SALD dataset has a longer age range (19–80 years) than that of the UK Biobank, we generalized the model to SALD subjects aged 45–80 years ($n = 271$). We showed that the XGBoost predictors could be successfully applied to this independent set with $r = 0.650$ ($CI = 0.576\text{--}0.714$), $MAE = 7.542$, and $RMSE = 8.707$ between predicted age and chronological age; after using the age-bias correcting model from the main experiment to remove the age-bias, the prediction performance...
Correlation between the corrected predicted age (i.e., BrainAge) and chronological smoking statuses showed different extents of brain ageing depending on BrainAge, and cognition using data from the UK Biobank. Using the XGBoost technique combined with a nested five-fold CV, we successfully significantly improved, with $r = 0.933$, CI = 0.576–0.714, MAE = 3.804, and RMSE = 4.490 (Fig. S5). Given that the SALD has less available demographic information than the UK Biobank, this result should be considered exploratory.

### 4. Discussion

The present study quantified the associations between smoking, BrainAge, and cognition using data from the UK Biobank. Using the XGBoost technique combined with a nested five-fold CV, we successfully predicted the BrainAge in either the control group or the smokers’ group with relatively high accuracy. We found that smokers with different smoking statuses showed different extents of brain ageing depending on smoking frequency and cessation. Smokers who currently smoked on most or all days had larger BrainAge Gaps than controls, while those who smoked occasionally did not differ significantly from controls. Higher lifetime exposure (i.e., pack-years) was associated with an older-appearing brain, while after quitting smoking, smokers showed a younger-appearing brain with a longer quitting duration. A larger BrainAge Gap was significantly associated with worse cognitive measures, and these associations were stronger if only smokers were considered. Smoking was also associated with selective cognitive functions, including reaction time and symbol digit substitution scores, with associations being partially mediated by brain age. This study’s robust findings will help scientists understand the association between smoking and the brain and cognition.

With the rapid development of computational methods for machine learning and publicly accessible large-scale neuroimaging databases, “brain age” analysis has been widely applied, which can reduce the dimensionality of neuroimaging data into a single dimension, providing comprehensible biomarkers of individual brain ageing. However, a uniform methodological approach for developing brain age predictors has not been established. Popular machine learning methods for brain age prediction include support vector regression, relevant vector regression, Gaussian process regression, and deep learning. Although the principles and algorithms are different, recent studies have shown that these methods have a close prediction performance (Baecker et al., 2021). We chose XGBoost because of its resource efficiency and

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The difference in BrainAge Gap between the smoker groups and the control group.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference (CI)</td>
</tr>
<tr>
<td>All Smoker - Control</td>
<td>0.304 (0.223, 0.384)</td>
</tr>
<tr>
<td>Current - Control</td>
<td>1.19 [0.978, 1.401]</td>
</tr>
<tr>
<td>Ex1 - Control</td>
<td>0.859 [0.393, 1.326]</td>
</tr>
<tr>
<td>Ex2 - Control</td>
<td>0.552 [0.445, 0.659]</td>
</tr>
<tr>
<td>Light1 - Control</td>
<td>0.478 [0.122, 0.834]</td>
</tr>
<tr>
<td>Light2 - Control</td>
<td>0.092 [-0.04, 0.224]</td>
</tr>
<tr>
<td>LightX - Control</td>
<td>-0.041 [-0.15, -0.007]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Associations of BrainAge Gap with Pack year and Quitting duration.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$ (SE)</td>
</tr>
<tr>
<td>Ever + Control</td>
<td>0.027 (0.002)</td>
</tr>
<tr>
<td>Current + Control</td>
<td>0.035 (0.004)</td>
</tr>
<tr>
<td>Not adjusted Pack year</td>
<td>-0.025 (0.003)</td>
</tr>
<tr>
<td>Adjusted Pack year</td>
<td>-0.015 (0.004)</td>
</tr>
</tbody>
</table>

| a | This analysis was restricted to Ever smoker + Control and Current smoker + Control, respectively. Ever smoker included current and Ex-smoker; |
| b | This analysis is divided into two cases: whether also to adjust for pack-year or not. |
XGBoost trained using data from the UK Biobank could not be extrapolated to children and adolescents (Chen and Guestrin, 2016), that is, they use a regression model to remove age dependence from brain age predictions. More explicitly, not controlling for the age-BrainAge relationship results in an approximately 5.2 years (Beheshti et al., 2019), which may provide an efficacious propaganda slogan to reduce tobacco consumption autonomously. Such associations of the smoking cessation–healthy index have also been found in cortical thickness (Karama et al., 2015), peripheral blood (Tsaprouni et al., 2014), respiratory organs (Wu et al., 2019), etc.

In spite of these findings, we could not identify a causal relationship between smoking and brain ageing. Two perspectives are presented here (Cole and Franke, 2017; Elliott et al., 2019). One perspective called the “geroscience perspective” (Kennedy et al., 2014) has suggested that all individuals have a BrainAge Gap close to zero at some point in early development, and the BrainAge Gap then diverges with time from chronological age, as genetic, environmental, and lifestyle factors create variations in the rate of brain ageing. Based on this viewpoint, our findings indicated that different smoking intensities and frequencies in smokers’ lifetime can cause different rates of brain ageing, and smoking cessation can slow this trend, thereby reducing the risk of age-related diseases. Another perspective is known as the “early system-integrity” (Deary, 2012) supposes that individuals vary in their brain health from the beginning of life and an older BrainAge at midlife reflects central nervous system variation that has been present since childhood and is stable for decades. From the system integrity perspective, BrainAge Gap is associated with smoking parameters because the BrainAge Gap is an indicator of compromised lifelong brain health, and individuals with larger BrainAge Gaps would likely be addicted to smoking, while those with relatively smaller BrainAge Gaps may find it easier to quit. In other words, the variation in brain ageing could, in part, be a precipitating factor for initiating smoking. Note that both perspectives are not mutually exclusive and help explain the association between smoking and accelerated brain ageing.

There are several limitations to this work. Firstly, since the results of demonstrated superior performance in previous machine learning competitions (Chen and Guerstrin, 2016). The prediction performance with MAE = approximately 4 years (RMSE = approximately 5.2 years) in the control or smoker groups reported in this study is slightly better than that reported in many recent larger-scale brain-age studies (e.g., MAE = 5.78 (de Lange et al., 2019), MAE = 5.11 (Beheresht et al., 2019)). MAE = 4.45 (Wang et al., 2019). In addition to using the advanced machine learning technique to improve the prediction power, an efficient method may be to cover a broader age range of the training data from early adulthood to older adulthood, allowing the model to acquire information about the relationship between imaging patterns and chronological age as much as possible. In our practice, we found that the XGBoost trained using data from the UK Biobank could not be extrapolated to data on younger people (e.g., 19–44 years) from the SALL dataset with a high degree of accuracy. In addition to the nature of the tree-based model, it may also be because the model lacks information on the relationship between brain and age in the younger population. It is more methodologically important to consider age bias (correlation between BrainAge and true age) in brain age predictions. More explicitly, not controlling for the age-BrainAge relationship results in an extensive set of spurious results: BrainAge Gap can be trivially correlated with cognitive variables if the latter are correlated with chronological age. Although many correction procedures have been proposed, many are mathematically or ideologically identical (de Lange and Cole, 2020), that is, they use a regression model to remove age dependence after the training model. Ideally and directly, this systematic bias of the estimated BrainAge Gap should be considered in the model training stage; this method is being developed (Treder et al., 2021).

Although smoking is associated with a larger BrainAge Gap (Bittner et al., 2021; Ning et al., 2020), we reported a more detailed relationship of the BrainAge Gap in different smoking statuses by comprehensively considering the smoking characteristics in a large-scale population. Although increased chronological age is relevant to the atrophic brain and a greater risk of developing neurodegenerative disorders such as dementia and cognitive decline (Abbott, 2011; Fjell and Walhovd, 2010), our findings may imply that smoking seems to reinforce this association, suggesting that smokers with a larger BrainAge Gap (i.e., an apparent-older brain) are more likely to develop age-related conditions (Debette et al., 2011; Durazzo et al., 2014). There was no significant association between the BrainAge Gap and light smoking history; nevertheless, we cannot simply translate this observation to the assumption that a previous light smoking history does not affect the brain. Perhaps this association was comparatively small and was offset by other healthy lifestyles after quitting smoking, such as physical exercise (Steffener et al., 2016). The magnitude of the BrainAge Gap in smokers is generally smaller than that in patients with neuropsychiatric diseases such as schizophrenia (+5.5 years) (Routoulou et al., 2014), Alzheimer’s disease (+5.76 years) (Low et al., 2016), and traumatic brain injury (+4.66 years) (Cole et al., 2015). Therefore, this association should not be overstated. In line with recent findings, we found that the more the participants smoked, the stronger the relationship between higher pack-years and bigger BrainAge Gaps, suggesting that this association was mostly driven by high lifetime smoking. We found a reverse trend of smoking cessation–BrainAge Gap association in the same population, which may provide an efficacious propaganda slogan to reduce tobacco consumption autonomously. Such associations of the smoking cessation–healthy index have also been found in cortical thickness (Karama et al., 2015), peripheral blood (Tsaprouni et al., 2014), respiratory organs (Wu et al., 2019), etc.

Table 4

<table>
<thead>
<tr>
<th>Reaction time</th>
<th>Symbol digit substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>p (SE)</td>
<td>t-value</td>
</tr>
<tr>
<td>Status</td>
<td>Current smoker</td>
</tr>
<tr>
<td></td>
<td>smoker (2.833)</td>
</tr>
<tr>
<td>Ex.</td>
<td>2.005</td>
</tr>
<tr>
<td></td>
<td>smoker1 (6.24)</td>
</tr>
<tr>
<td>Ex.</td>
<td>−2.917</td>
</tr>
<tr>
<td></td>
<td>smoker2 (1.451)</td>
</tr>
<tr>
<td>Light</td>
<td>1.334</td>
</tr>
<tr>
<td></td>
<td>smoker1 (4.763)</td>
</tr>
<tr>
<td>Light</td>
<td>1.824</td>
</tr>
<tr>
<td></td>
<td>smoker2 (1.763)</td>
</tr>
<tr>
<td>Light</td>
<td>1.015</td>
</tr>
<tr>
<td></td>
<td>smokerX (1.45)</td>
</tr>
<tr>
<td>Pack-year</td>
<td>0.350</td>
</tr>
<tr>
<td>Quitting duration</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th>Association of BrainAge Gap with 5 cognitive measures.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reaction time</td>
</tr>
<tr>
<td>Fluid intelligence</td>
</tr>
<tr>
<td>Numeric memory</td>
</tr>
<tr>
<td>Pairs matching</td>
</tr>
<tr>
<td>Symbol digit substitution</td>
</tr>
</tbody>
</table>
this study were based on cross-sectional data, as discussed above, it is impossible to disentangle the causality of these associations between smoking and brain ageing differences identified here. Future longitudinal studies are required to determine the causality. Secondly, the brain-age model in this study solely used information from structural MRI data, combining other modalities (e.g., diffusion MRI and functional studies) are required to determine the causality. Secondly, the smoking parameters data was recalled, which could result in a recall bias. Thirdly, the smoking variables data was obtained from each subject. All UK Biobank data used in this work were obtained under Data Access Application 19542 and are available to eligible researchers under UKB application 19542 (PI Jianfeng Feng).

Data availability

All UK Biobank data used in this work were obtained under Data Access Application 19542 and are available to eligible researchers through the UK Biobank (www.biobank.ac.uk).

Declararion of Competing Interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pnpbp.2021.110471.

Acknowledgments

The authors would like to thank all contributors to the UK Biobank Project. The authors would like to thank all contributors to the UK Biobank Project. The authors would like to thank all contributors to the UK Biobank Project. The authors would like to thank all contributors to the UK Biobank Project. The authors would like to thank all contributors to the UK Biobank Project. The authors would like to thank all contributors to the UK Biobank Project.

References


Z. Lini et al. in Progress in Neuropsychopharmacology & Biological Psychiatry 113 (2022) 110471