nature portfolio

Corresponding author(s): Jianfeng Feng
Wei Cheng
Jintai Yu

Last updated by author(s): Jul 26, 2023

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was involved in data collection (data used is all directly available from UK Biobank, as described in detail in the paper ) |
| Data analysis | Freesurfer v6.0 was used to process the imaging data.<br>PLINK 2.0 and PRSice version 2 (www.PRSice.info) were used to perform genome-wide association analysis and calculate the polygenic risk score respectively.<br>R version 4.2.1 package:<br>survival 3.4-0 was used to perform Cox proportional hazard regression model,<br>TwoSampleMR 0.5.6 and RadialMR 1.0 was used to perform Mendelian Randomization analysis,<br>lavaan 0.8 was used to perform structural equation model.<br>Scripts used to perform the analyses are available at https://github.com/yjzhao1004/lifestyle_depression |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Access to individual-level UK Biobank data, both phenotypic, neuroimaging and genotype, is available to bona fide researchers through application on the UK Biobank website (https://www.ukbiobank.ac.uk). Additional information about registration for access to the data is available at http://www.ukbiobank.ac.uk/register-apply/. Use of UK Biobank data was performed under application number 19542. Summary statistics from previous GWAS of depression which were used in this study are publicly available through the Psychiatric Genomics Consortium (PGC – Psychiatric Genomics Consortium (unc.edu)) and can be downloaded from https://pgc.unc.edu/for-researchers/download-results/.

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | Both male and female subjects from UK biobank study were included in the study. Summary statistics on sex distributions were reported in Table 1. All statistic models were adjusted for sex. |
| Population characteristics | We utilized a total of 502,409 participants from UK Biobank cohort to perform survival analysis, correlation analysis, and structural equation model. For survival analyses, we retrieved 287,282 individuals (mean age 57.52, with 50.70% female) who participated in the assessment of lifestyle factors at baseline (2006-2010) and had linkage to clinical diagnosis records. The number of participants in healthy or unhealthy lifestyle in each factor was listed in Table 1. The mean score of 7 lifestyle was 4.75 (s.d., 1.36), of which 1.25% were categorized as following an unfavorable lifestyle (scores ranging from 0-1), 38.90% followed an intermediate lifestyle (scores ranging from 2-4), and 59.85% followed a favorable lifestyle (scores ranging from 5-7). Of the 287,282 participants, 12,916 of them had an onset of depression during a median follow-up of 9.01 years. Demographic characteristics of the participants and distribution of lifestyle factors were provided in Table 1. The polygenic risk scores for depression from 197,344 participants (mean age 57.81, with 49.82% female) were normally distributed (see Supplementary Fig. 5) and categorized into three levels, low risk (25.09%), intermediate risk (49.97%), high risk (24.94%). Brain structural imaging data were collected in 2014, containing T1-weighted structural MRI from 32,839 participants used in the correlation analysis. Blood chemistry and cell count indicators from about 480,000 participants and NMR metabolic biomarkers from about 120,000 participants were collected at baseline. The sample size of each biomarker in correlation analyses were presented in Supplementary table 18 and 19. We additionally used data from a total of 448,849 participants with completed Patient Health Questionnaire-4 (2006-2010) score. Combining the data of lifestyle score, PRS, immunometabolic markers, brain structural imaging and depression, the final analysis included 18,244 participants (mean age 55.20, with 48.79% female) to estimate the structural equation model. Fig. 1 showed the research guideline of the study. Supplementary Fig.1 and Supplementary table 4 presented the sample size utilized in each analysis. |
| Recruitment | The UK Biobank is a prospective, population-based cohort that recruited more than 500,000 participants aged 37 - 73 years who attended 1 of 22 assessment centers across the United Kingdom between 2006 and 2010. Previous investigation showed UK biobank subject to a healthy sample bias. |
| Ethics oversight | The UKB has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). All participants provided informed consent through electronic signature at the baseline assessment. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No statistical methods were used to predetermine sample sizes and all currently available sample in the UK Biobank were used including 394,053 subjects with lifestyle factor data, 287,282 subjects with depression diagnosis record, 32,839 subjects with imaging data, about 480,000 subjects with blood chemistry data, about 120,000 subjects with metabolic markers data and 337,151 subjects with genotype data. The sample sizes were provided in the method section. |

| | |
|---|---|
| Data exclusions | For cox proportional hazard regression model, we excluded patients with missing lifestyle factor data, baseline depression, unavailable baseline data, missing follow-up data and those with self-report depression. Flow chart of the sample size selection in survival analyses and structural equation model were presented in Supplementary Fig.1.<br>For genome-wide association analyses on lifestyle score, we excluded single-nucleotide polymorphisms (SNPs) with call rates < 95%, minor allele frequency < 0.1%, deviation from the Hardy–Weinberg equilibrium with p < 1E-10 and selected subjects that were estimated to have recent British ancestry and have no more than ten putative third-degree relatives in the kinship table, consistent with the previous study. |
| Replication | All available data were used to maximize statistical power of the analysis therefore we did not repeat the analysis. |
| Randomization | All models were adjusted for age at baseline, sex, Townsend deprivation index, body mass index (BMI) and education levels. Association analyses involving neuroimaging data were additionally adjusted neuroimaging scanning sites and estimated total intracranial volume. Association analyses involving genetic data, PRS components (first 20 principal components of ancestry) were further added as covariates. Full details of covariates utilized in the statistical analyses see Supplementary Fig.2. |
| Blinding | Blinding was not applicable to this study as this study is observational. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☐ | ☒ MRI-based neuroimaging |

# Magnetic resonance imaging

## Experimental design

| | |
|---|---|
| Design type | Structural MRI |
| Design specifications | UK Biobank designed the imaging acquisition protocals including 6 modalities, covering structural, diffusion and functional imaging. The collection order is T1-weighted structural image, resting-state functional MRI, task functional MRI, T2-weighted FLAIR structural image, Diffution MRI and susceptibility-weighted imaging. T1-weighted structural image was acquired using straight sagittal orientation for 5 minutes. |
| Behavioral performance measures | Lifestyle assessments were first administered via touchscreen interface in the UK Biobank assessment center at the baseline visit (2006-2010). Seven aspects of lifestyle factor assessments including alcohol consumption, diet, physical activity, sleep duration, smoke, sedentary behavior and social connection were utilized to calculate lifestyle score in the current study (n = 394,053). Measurement of depressive symptoms via 4-item Patient Health Questionnaire-4 (PHQ-4) was first assessed in the UK Biobank assessment center (2006-2010, n=448,849). |

## Acquisition

| | |
|---|---|
| Imaging type(s) | T1-weighted structural imaging |
| Field strength | 3T |
| Sequence & imaging parameters | The EPI-based acquisitions utilize simultaneous multi-slice (multiband) acceleration. Biobank uses pulse sequences and reconstruction code from the Center for Magnetic Resonance Research (CMRR), University of Minnesota https://www.cmrr.umn.edu/multiband. The resolution is 1x1x1 mm and field of view is 208x256x256 matrix. Straight sagittal orientation is used. TR and TE are 2000ms and 2.01ms respectively. The flip angle is 8 deg. Detailed sequence and imaging parameters are openly available here: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/ brain_mri.pdf |
| Area of acquisition | Whole brain |

Diffusion MRI ☐ Used ☒ Not used

## Preprocessing

Preprocessing software: T1 images were processed with FreeSurfer; surface templates were used to extract imaging-derived phenotypes referred to as atlas regions' surface volume76. Subcortical regions were extracted via FreeSurfer's aseg tool. FreeSurfer aparc (Category ID 192) and aseg (Category ID 190) atlas corresponding to 68 cortical regions76 and 14 subcortical regions were used in this study. We used the Qoala-T approach to check FreeSurfer outputs and excluded those failed to pass quality control from the FreeSurfer imaging-derived phenotypes.

Normalization: see above

Normalization template: fsaverage

Noise and artifact removal: see above

Volume censoring: see above

## Statistical modeling & inference

Model type and settings: 1) Cox proportional hazard regression model was used to examine the association of lifestyle with risk of depression. 2) Mendelian randomization (MR) analysis was used to examine the causal relationship between lifestyle and depression. 3) Structural equation model was used to analyze the association between lifestyle, polygenic risk score, brain structure, immunometabolic function and depression.

Effect(s) tested: For cox proportional hazard regression model, MR analysis and structural equation model, Wald tests were utilized to derive the two-sided p value. Additionally, for survival analysis, the proportional hazards assumption was assessed using the Schoenfeld residuals method and satisfied. For MR analysis, we used inverse variance weighted method, simple median method, weighted median method and weighted mode method to estimate the causal effect.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016): voxel-wise association, voxel-wise Bonferroni correction

Correction: Bonferroni

## Models & analysis

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Functional and/or effective connectivity |
| ☒ | ☐ Graph analysis |
| ☒ | ☐ Multivariate modeling or predictive analysis |