nature portfolio

Corresponding author(s): Tabea Schoeler

Last updated by author(s): Sep 29, 2022

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 *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 *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	This research has been conducted with the UK Biobank Resource under application number 16389.
Data analysis	We used the following software to conduct the analyses: LD Score Regression as implemented in GenomicSEM (https://github.com/GenomicSEM/GenomicSEM), LDAK (http://dougspeed.com/downloads/), TwoSampleMR (https://mrcieu.github.io/TwoSampleMR/). All analytical scripts are available at https://github.com/TabeaSchoeler/TS2021_UKBBweighting

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

Data availability: Standard and probability weighted UK Biobank association statistics, computed using LDAK version 5.2, will be made available through the GWAS catalog.

Human research participants

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

Reporting on sex and gender	We used self-reported sex (biological attribute) in our study.
Population characteristics	In genome-wide analyses, we included UK Biobank participants of European ancestry passing standard GWA analysis quality control measures. All analyses were adjusted for batch, principal components (PC1-PC5), age and sex. Exclusions during QC process (phenotypic and genetic) are detailed in the Methods. Demographic information about the sample is provided in Supplementary Table 3.
Recruitment	The UK Biobank (UKBB) is a prospective population-based research resource focusing on the role of genetic, environmental and lifestyle factors in health outcomes in middle age and later life. More than 9,000,000 men and women between 40 and 69 registered with the UK NHS were invited to take part. Of those, 5.4% (~500,000 individuals) were recruited in 22 assessment centres across England, Wales and Scotland between 2006 and 2010.
Ethics oversight	The UK Biobank resource was approved by the UK Biobank Research Ethics Committee and all participants provided written informed consent to participate.

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	We conducted inverse probability weighted genome-wide association analyses (Neffective=94,643 – 102,215) and standard GWA (N=263,464 – 283,749) in UKBB participants selected for genome-wide analyses (UK Biobank participants of European ancestry passing standard GWA analysis quality control measures). Quality control filters for genome-wide analyses were applied to select participants (i.e., exclusion of related individuals, exclusion of non-White British ancestry based on principal components, high missing rate and high heterozygosity on autosomes) and genetic variants (Hardy–Weinberg disequilibrium P>1×10−6, minor allele frequency>1% and call rate>90%).
Data exclusions	We filtered the sample according to geographical region (excluding individuals from Scotland and Wales) to match the geographical regions included in the reference sample (HSE), and removed individuals with missing data in auxiliary variables used to generate the propensity scores.
Replication	We used the UK Biobank as it is the currently largest sample where participation bias correction through inverse weighted genome-wide association analyses can be performed. Our findings replicate previous genome-wide findings and highlight the extend to which these findings may be biased by selective participation. We did not select an independent replication sample. There are no genotype datasets of similar size in the UK for which sampling weights could be computed, making replication currently not feasible.
Randomization	Not applicable
Blinding	Not applicable

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