

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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 used for data collection.
Data analysis	PLINK 2.0 was used to run the genome-wide association studies (GWAS). REGENIE (v4.1) and METAL (version 2020-05-05) were used for the GWAS meta-analysis. For Mendelian randomization analyses, R (v4.3.1) was used with the TwoSampleMR (v0.5.11), MendelianRandomization (v0.9.0), MRlap (v0.0.3.2), and MVMR (v0.1) packages. The code used for the analyses has been made openly available on GitHub (see https://github.com/han09090/Estradiol_Brain_Mental_Health and https://doi.org/10.5281/zenodo.17099458).

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](#)

The summary statistics from the female-specific and male-specific GWAS on brain age gap and the female-specific GWAS on estradiol levels using the continuous

approach and the binary approach are openly available on the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>; study accessions: brain age gap in females: GCST90483374; brain age gap in males: GCST90483373; estradiol levels using the continuous approach: GCST90483375; estradiol levels using binary approach: GCST90483376). Data used for the GWAS are available from the UK Biobank (<https://www.ukbiobank.ac.uk>) and the Norwegian Mother and Child Cohort Study (<https://www.fhi.no/moba-en>). All other summary statistics included in the present study are openly available for download from the respective repositories or were obtained by contacting the authors, as described in the manuscript.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Due to the genetic nature of the present study, our analyses focused on biological sex. For GWAS conducted in the present study, sex was determined according to genetic sex (XX or XY chromosomes). Main analyses were conducted in females, as variables of interest included estradiol levels (the predominant sex hormone in females), age at menarche, age at menopause, length of reproductive span (age at menopause minus age at menarche), number of childbirth, oral contraceptive use, hormone replacement therapy use, history of hysterectomy, and history of oophorectomy. Sensitivity analyses were conducted in males for estradiol levels.

Reporting on race, ethnicity, or other socially relevant groupings

Due to the genetic nature of the present study, samples were limited to White European ancestry. Unfortunately, no large-scale databases were available to replicate the results in samples with other ancestries. For GWAS conducted in the present study, ancestry was determined according to genetic grouping as reported in the UK Biobank.

Population characteristics

Age ranges of the participants included in the GWAS were as follows:
continuous estradiol levels: 40.16 - 70.17 years (M = 48.30, SD = 5.96)
binary estradiol levels: 40.16 - 71.10 years (M = 57.39, SD = 7.93)
brain age gap (female sample): 45.13 - 81.83 years (M = 63.56, SD = 7.27)
brain age gap (male sample): 46.07 - 82.18 (M = 64.82, SD = 7.57)

Recruitment

No participants were recruited specifically for the present study. However, for large-scale databases as used in the present study (e.g., UK Biobank), healthy volunteer biases and survivor biases may influence the findings. The present study investigated estradiol levels using data from the UK Biobank. This may have influenced the results, as estradiol levels are generally lower in middle-aged to older females. However, the present study conducted several sensitivity analyses and replicated the results in separate samples of pre- and postmenopausal females as well as males and found consistent patterns.

Ethics oversight

The UK Biobank has received ethics approval from the National Health Service National Research Ethics Service (ref 11/NW/0382). The study dataset complies with the Helsinki Declaration, with informed consent obtained from all participants. This research has been conducted using the UK Biobank under Application 27412. MoBa is regulated by the Norwegian Health Registry Act. The present study was approved by the Regional Committees for Medical and Health Research Ethics (2016/1226/REK).

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](https://www.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

Sample size was determined according to largest possible sample size for available GWAS summary statistics, while avoiding sample overlap (when possible) between exposure and outcome samples for Mendelian randomization analyses. For the GWAS conducted in the present study, sample size was determined by availability of data. To increase sample size for estradiol levels, sensitivity analyses were conducted using a binary approach for this variable.

Data exclusions

For all GWAS conducted in the present study, participants that were not of White European ancestry were excluded. For continuous estradiol levels, males were excluded, and females were excluded if they had levels under the detection limit or if they responded "prefer not to answer" or "do not know/not sure" on any of the covariates (i.e., menopausal status, oral contraceptive use, hormone replacement therapy use, history of hysterectomy, and history of bilateral oophorectomy). For brain age gap, participants who did not have both T1-weighted and diffusion-weighted magnetic resonance imaging data, outliers based on Euler numbers more than 4 SD above the mean, as well as participants with ICD-10 diagnoses known to impact brain health (ICD field F including F00-03 for Alzheimer's disease and dementia and F06.7 for mild cognitive disorders, and excluding depressive disorders), diseases of the nervous system (ICD field G including inflammatory and neurodegenerative diseases, except G55-59), or stroke (ICD field code I64) were excluded. For brain age gap in males, females were excluded. For brain age gap in females, males were excluded. Exclusion criteria were pre-established for brain age gap analyses. For the GWAS meta-analysis, only males were included. Cases were defined as having a lifetime (primary or secondary for UK Biobank;

following the Norwegian Patient Registry for MoBa) diagnosis of a depressive episode or recurrent depressive episode, according to the ICD-10 (F32 or F33). Cases with a lifetime (primary or secondary) diagnosis of schizophrenia, schizotypal, or delusional disorder (F20-F29), mania or bipolar disorder (F30 or F31), or personality disorder (F40 or F61) were excluded. Further, controls with a lifetime (primary or secondary) diagnosis of any mood disorder (F30-F39), schizophrenia, schizotypal, or delusional disorder (F20-F29), or personality disorder (F40 or F61) were excluded.

Replication

We replicated the results on estradiol levels in several samples: a sample of premenopausal-only females from the UK Biobank, a sample of postmenopausal-only females from the UK Biobank, an independent sample of postmenopausal females from the LIFE-Adult and LIFE-Heart studies, and male-only samples. Further, results on continuous estradiol levels were replicated using binary estradiol levels (above or below detection limit). For analyses with depression as an outcome with sample overlap, analyses were replicated using a subsample of the depression sample excluding the UK Biobank. Results were consistently non-significant across samples and estradiol phenotypes.

Randomization

Randomization was not relevant for the present study.

Blinding

Blinding was not used in the present study.

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
- ☒ ☐ Plants

Methods

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

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a

Magnetic resonance imaging

Experimental design

Design type

Resting state

Design specifications

n/a

Behavioral performance measures

n/a

Acquisition

Imaging type(s)

T1-weighted and diffusion-weighted images

Field strength

See UK Biobank for details

Sequence & imaging parameters

See UK Biobank for details

Area of acquisition

Whole brain

Diffusion MRI ☒ Used ☐ Not usedParameters *Specify # of directions, b-values, whether single shell or multi-shell, and if cardiac gating was used.*

Preprocessing

Preprocessing software	FreeSurfer v5.3 for raw T1-weighted images. Diffusion-weighted images were processed using an optimized diffusion pipeline.
Normalization	Grey matter and white matter MRI features were derived from T1-weighted and diffusion-weighted images, respectively, and used for multimodal brain age prediction (see Alfaro-Almagro et al., 2018 and Miller et al., 2016 for details on MRI data acquisition and protocols in the UKB). Automated surface-based morphometry and subcortical segmentation pipelines in FreeSurfer v5.3 (Fischl, 2012) were used to process raw T1-weighted images. The standard set of FreeSurfer-derived subcortical and cortical summary statistics (Fischl, 2012) were further supplemented by a fine-grained cortical parcellation scheme (Glasser et al., 2016) to extract cortical thickness, area, and volume for 180 regions of interest per hemisphere. Diffusion-weighted data were processed using an optimized diffusion pipeline (Korbmacher et al., 2024; Maximov et al., 2019; Maximov et al., 2021) and similarly controlled for scanner site using linear models. The diffusion-weighted data passed TBSS post-processing quality control through the YTTRIUM algorithm (Korbmacher et al., 2024; Maximov et al., 2019; Maximov et al., 2021).
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	Outliers based on Euler numbers more than 4 SD above the mean were excluded.
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

Statistical modeling & inference

Model type and settings	MRI data was used for brain age estimation. Brain age was computed using the XGBoost (eXtreme Gradient Boosting; Chen & Guestrin, 2016) regression model, based on a decision-tree ensemble algorithm. Hyperparameters were tuned using a nested cross-validation with 5 inner folds for randomized search and 10 outer folds for validation of the model (see general model setup: https://github.com/amdelange/brainage_women/blob/main/python/Brain_age_prediction.py and https://xgboost.readthedocs.io/en/latest/python/index.html)
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
(See Eklund et al. 2016)	
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

Models & analysis

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input checked="" type="checkbox"/> Multivariate modeling or predictive analysis
Multivariate modeling and predictive analysis	T1-weighted features included cortical thickness, volume area, and summary statistics. Diffusion-weighted features included diffusion tensor imaging, spherical mean technique, diffusion kurtosis imaging, white matter tract integrity, and white matter features based on Johns Hopkins University atlases for white matter tracts and labels (with 0 thresholding) for region of interest and mean values. T1-weighted and diffusion-weighted features were used for brain age prediction. Brain age was computed using the XGBoost (eXtreme Gradient Boosting; Chen & Guestrin, 2016) regression model, based on a decision-tree ensemble algorithm. Hyperparameters were tuned using a nested cross-validation with 5 inner folds for randomized search and 10 outer folds for validation of the model (see general model setup: https://github.com/amdelange/brainage_women/blob/main/python/Brain_age_prediction.py and https://xgboost.readthedocs.io/en/latest/python/index.html)