

No causal links between estradiol and female's brain and mental health using Mendelian randomization

Received: 31 July 2024

Accepted: 24 October 2025

Published online: 05 December 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

The role of estradiol in depression and Alzheimer's disease – brain disorders that disproportionately affect females – is debated. Results from observational studies are inconsistent and limited by confounding and reverse causation. To overcome these limitations, we perform two-sample Mendelian randomization. We run genome-wide association studies on sex-specific brain age gap, a proxy of brain health, and female-specific estradiol levels using data from the UK Biobank. We test for causal links between genetically-predicted factors related to estradiol exposure (estradiol levels in pre- and postmenopausal samples, reproductive span, age at menarche, age at menopause, number of childbirths) and brain age gap, Alzheimer's disease and depression as outcomes. We replicate our analyses on estradiol levels in males. Here, we find no significant associations between estradiol exposure and brain health across samples and robust methods, indicating an absence of constant causal effects and suggesting that hormonal fluctuations may drive links between estradiol and brain health.

Females are at a significantly higher risk of depression with higher chronicity¹ and late-onset Alzheimer's disease with faster progression², relative to males. Estradiol, the predominant estrogen in the female body, is a modulator of brain structure and function, and has repeatedly been implicated in the development of depression and Alzheimer's disease in females³.

The estrogen hypothesis postulates that estrogens exert a neuroprotective effect on the brain and cognitive functions⁴. Aligning with this hypothesis, a range of factors indicative of higher lifetime estradiol exposure, such as an older age at menopause and a longer reproductive span (age at menopause minus age at menarche), have been linked to lower risk for depression⁵ and dementia⁶. Further, these factors have been associated with markers of brain health using data from the UK Biobank (UKB)^{7–9}, a prospective population-based cohort from the United Kingdom (see Sudlow et al.,¹⁰ for details on the UKB). For instance, a lower brain age gap has been linked to a higher number of childbirths⁷, an older age at natural menopause⁸, and a longer reproductive span⁹. Brain age gap is defined as the difference between predicted brain age derived from magnetic resonance imaging (MRI) measures and chronological age¹¹. A higher brain age gap has been

linked to a range of negative health outcomes, including Alzheimer's disease and depression^{12,13}. However, some observational studies contradict the estrogen hypothesis, linking a longer reproductive span, an earlier age at menarche, and a later age at menopause with smaller total brain volume¹⁴. Inconclusive findings on the relationship between estradiol and brain and mental health might partly be traced back to differences in the data used, the type of hormone exposure (endogenous vs. exogenous) assessed, and measurements or proxies used to ascertain estradiol exposure.

Further, the predominant use of observational study designs encompasses challenges, such as confounding and reverse causation, that might contribute to inconsistent findings across studies¹⁵. To overcome some of these limitations, recent studies have implemented Mendelian randomization, which uses genetic variants to test for causal exposure-outcome relationships¹⁵. The idea behind Mendelian randomization mirrors that of a randomised controlled trial, however, instead of randomising individuals into experimental groups, the randomization occurs through the assignment of genetic variants at conception¹⁵. Two-sample Mendelian randomization, where variant-exposure and variant-outcome associations are taken from different

✉ e-mail: hannah.oppenheimer@psykologi.uio.no; claudia.barth@charite.de

samples derived from the same underlying population, has become increasingly popular, as it makes use of summary statistics from genome-wide association studies (GWAS) and increases statistical power¹⁵.

Using this approach, previous studies have examined relationships between variables related to estradiol as exposures, including age at menarche, age at menopause, and estradiol levels, and outcomes related to brain and mental health, including Alzheimer's disease and depression risk, largely reporting no causal associations or small effect sizes^{16–22}. However, many previous studies have used small or mixed-sex samples, likely as a consequence of the lack of available large-scale, sex-stratified GWAS. As previous studies have highlighted different genetic effects regarding estradiol levels between females and males²³, sex-specific effects might be overlooked and obtained estimates may be biased when using mixed-sex samples. Furthermore, many other factors related to lifetime estradiol exposure commonly used in observational studies, such as reproductive span and number of childbirths, as well as other indicators of brain health, including brain age gap, have not yet been examined using Mendelian randomization. Thus, a comprehensive approach investigating these factors in female-only samples has so far been lacking. Such an approach can improve our mechanistic understanding of sex differences in mental disorders and neurodegenerative diseases. Based on previous observational studies, we hypothesised that higher estradiol levels and factors related to higher lifetime estradiol exposure are associated with a lower brain age gap (i.e., younger brain age relative to chronological age), a lower risk for Alzheimer's disease, and a lower risk for depression, thus showing protective effects of estradiol on brain health and mental health in females.

In this preregistered study, we conduct two-sample Mendelian randomization analyses in female-only samples and find no causal effects of lifetime estradiol exposure on the included proxies of brain and mental health. In addition to openly available sex-stratified GWAS, we use data from the UKB to run and annotate a female-specific GWAS on brain age gap as well as a female-specific GWAS on estradiol levels using a continuous and a binary approach (i.e., above or below detection limit). As exposure variables, we use genetically-predicted estradiol levels as well as factors related to lifetime estradiol exposure (i.e., reproductive span, age at menarche, age at menopause, and number of childbirths) and examine each relationship with genetically-predicted brain age gap, risk for Alzheimer's disease, and risk for depression as outcome variables. Due to methodological challenges of measuring estradiol levels in pre- and postmenopausal females (e.g., effects of menstrual cycle and menopausal status), likely impacting GWAS results²⁴, we conduct our analyses on continuous estradiol levels in combined as well as stratified samples of pre- and postmenopausal females from the UKB and use the binary approach for sensitivity analyses. To further ensure robustness of our results across samples, we replicate our findings in an independent sample of postmenopausal females and, as estradiol levels are known to fluctuate less in males, in a male-only sample. To this end, we run a male-specific GWAS on brain age gap using data from the UKB and a GWAS meta-analysis including data from the UKB, the Norwegian Mother and Child Cohort Study (MoBa; a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health; see Magnus et al.²⁵ for details), and a previously conducted GWAS. We run supplementary analyses using factors related to exogenous hormone use and health-related procedures that likely impact estradiol levels (i.e., oral contraceptive use, hormone replacement therapy (HRT) use, history of hysterectomy, and history of oophorectomy) as exposures. As further sensitivity analyses, we use risk for recurrent depression as an outcome to assess potential influences of disease burden, as well as a subsample of the depression GWAS to avoid sample overlap. We conduct multivariable Mendelian randomization, including body mass index (BMI) as an exposure variable, to account for potential

pleiotropic effects, as BMI is known to be associated with estradiol exposure and brain and mental health^{16,19,23}. For the Mendelian randomization analyses, we apply the inverse-variance weighted (IVW) method and robust estimation methods, including MR-Egger, weighted median, simple mode, and weighted mode, as well as MRlap for analyses with sample overlap.

Results

GWAS of brain age gap in females

We identified two independent genomic loci, on chromosomes 5 and 17, that were significantly associated with brain age gap in females (see Fig. 1 for Manhattan plot and Supplementary Fig. 4 for QQ-plot). The loci included 3 lead SNPs and 4 independent SNPs and mapped onto 14 genes (Supplementary Table 5 and Supplementary Fig. 9). Out of the 14 genes, 5 genes (*MAPT*, *NSF*, *ARHGAP27*, *KANSL1*, and *PLEKHMI*) expressed higher, on average, across brain tissue compared to the other genes. The other nine genes largely showed lower average expression levels across tissue types (Supplementary Fig. 12). The gene-set enrichment analysis identified 1 positional gene set and 28 associated phenotypes (Supplementary Fig. 15). We identified no significant genomic loci in the GWAS conducted on brain age gap in the male-only sample (see Supplementary Note 4).

GWAS of estradiol levels in females

For the GWAS conducted on continuous estradiol levels in a combined sample of pre- and postmenopausal females, we identified two independent genomic loci, on chromosomes 17 and 19, that were significantly associated with estradiol levels (see Fig. 2 for the Manhattan plot and Supplementary Fig. 5 for the QQ-plot). The loci included 3 lead SNPs and 4 independent SNPs and mapped onto 21 genes (Supplementary Table 6 and Supplementary Fig. 10). A cluster of genes (*SHBG* and *SLC35G6*) expressed lower, on average, across tissue types, compared to the other genes (Supplementary Fig. 13). The gene-set enrichment analysis identified 1 positional gene set and 10 associated phenotypes (Supplementary Fig. 16). We identified nine independent genomic loci in the GWAS conducted on binary estradiol levels (see Supplementary Note 6 for details).

Mendelian randomization analyses

No significant causal relationships were found between estradiol levels, using the continuous measures as the exposure, and brain age gap, Alzheimer's disease, and depression as outcomes (Table 1). The results were consistently non-significant across the combined, premenopausal-only, and postmenopausal-only samples from the UKB. Similarly, the replication analyses in the independent postmenopausal sample from the LIFE studies were non-significant (Table 1). Across the samples used for continuous estradiol levels, few instrumental variables were identified at the genome-wide significance threshold, therefore, a lower threshold ($p < 5 \times 10^{-6}$) was applied for selecting instrumental variables (see Supplementary Note 3 and Supplementary Table 4 for details). All results remained non-significant across the estimation methods, except for the analysis with estradiol levels in the combined pre- and postmenopausal sample as an exposure and Alzheimer's disease as an outcome, which was significant using the MR-Egger method ($b = 1.27$, $se = 0.48$, $p = .02$; Supplementary Table 9).

The sensitivity analyses using binary estradiol levels for the combined pre- and postmenopausal, premenopausal-only, and postmenopausal-only females from the UKB were not significant for any of the outcomes (Supplementary Table 10). For the male samples, no significant associations were found between estradiol levels and brain age gap, Alzheimer's disease, and depression (Supplementary Table 10). The results remained non-significant across the estimation methods.

No significant associations were found between the exposures reproductive span, age at menopause, and number of childbirths and

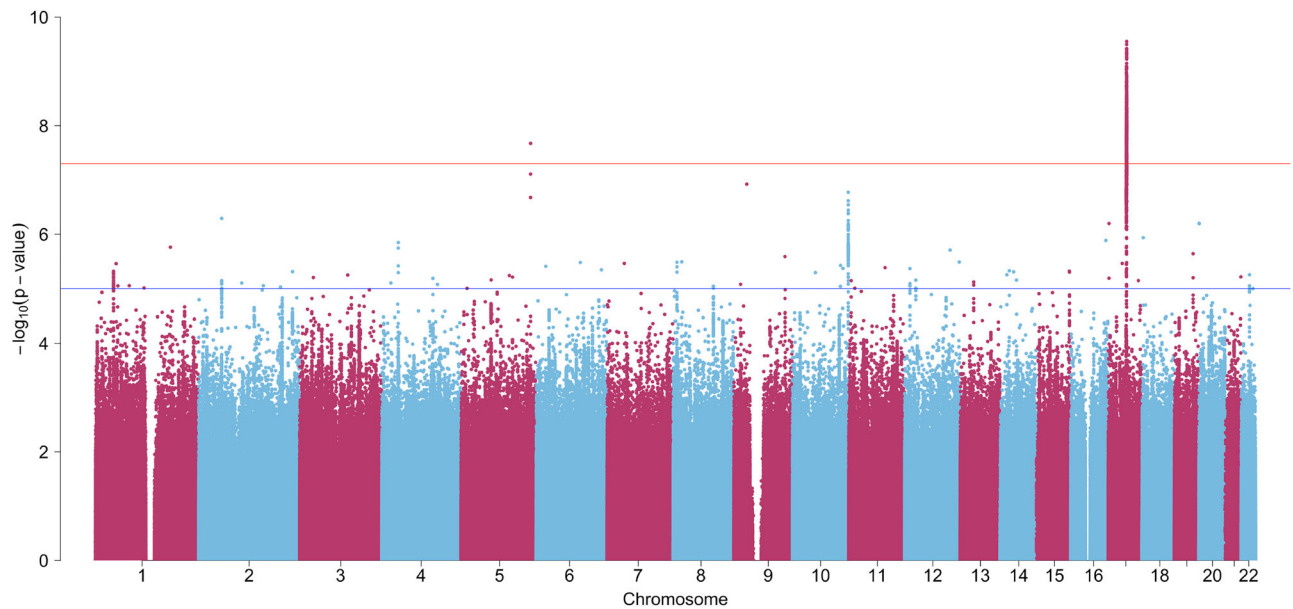


Fig. 1 | Manhattan plot of brain age gap in females. The blue/lower and red/upper lines indicate the suggestive ($p < 5 \times 10^{-5}$) and the genome-wide significance ($p < 5 \times 10^{-8}$; i.e., accounting for multiple comparisons) thresholds, respectively. The genome-wide association study was run using the standard additive model of linear associations. Source data are provided as a Source Data file.

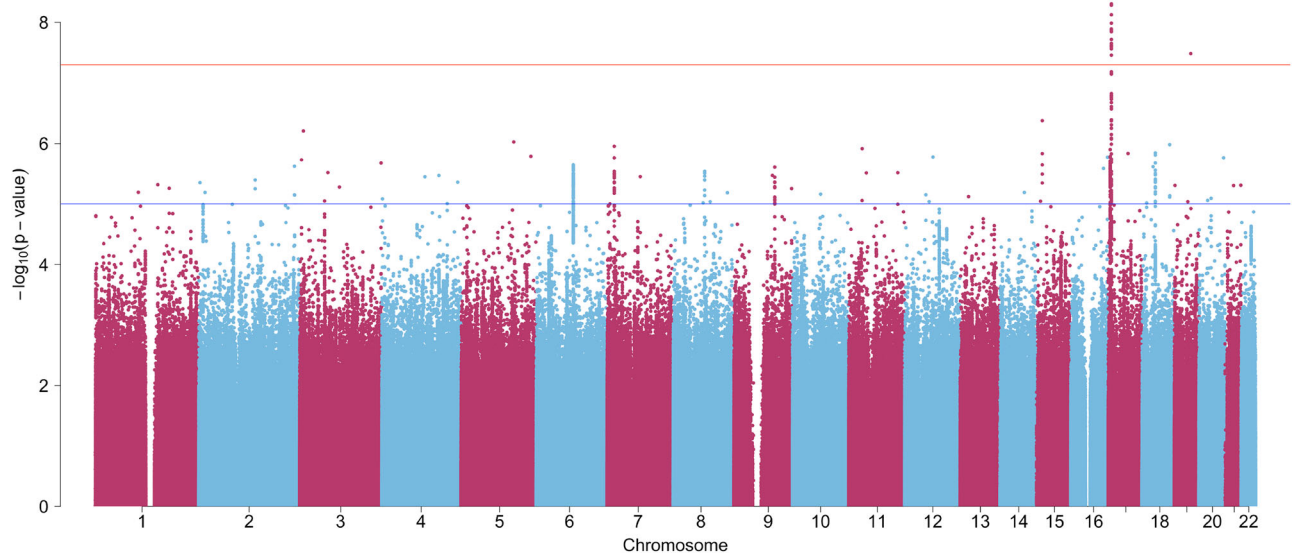


Fig. 2 | Manhattan plot of estradiol levels (continuous approach). The blue/lower and red/upper lines indicate the suggestive ($p < 5 \times 10^{-5}$) and the genome-wide significance ($p < 5 \times 10^{-8}$; i.e., accounting for multiple comparisons) thresholds, respectively. The genome-wide association study was run using the standard additive model of linear associations. Source data are provided as a Source Data file.

the outcomes brain age gap, Alzheimer's disease, and depression (Table 2). Consistently, no significant effects were found in the supplementary analyses using oral contraceptive use, HRT use, history of hysterectomy, and history of oophorectomy as exposure variables (Supplementary Note 7 and Supplementary Table 8). All results remained robust across the estimation methods. Similarly, age at natural (non-surgical) menopause as an exposure was not significant in the sensitivity analysis for Alzheimer's disease as an outcome (Supplementary Table 11). A significant association was found for age at menarche with depression as an outcome ($b = -0.09$, $se = 0.04$, $p = .04$), however, this result did not remain significant after adjusting for multiple comparisons ($p_{FDR} = .76$) and was not significant when using any of the other estimation methods (Fig. 3). For age at menarche as an exposure, no significant relationships were found with

brain age gap or Alzheimer's disease as outcomes (Table 2). Results were robust across the estimation methods for Alzheimer's disease and largely for brain age gap, except for the weighted median which was significant ($b = -0.44$, $se = 0.21$, $p = .04$; Supplementary Table 9).

When using MRlap to correct for potential biases resulting from sample overlap, the results of the analyses with overlapping exposure and outcome samples (where both samples included data from the UKB) remained robust, with no significant differences between the IVW estimates and the corrected estimates (Supplementary Table 9). Further, the sensitivity analyses using the depression subsample excluding UKB for analyses with sample overlap replicated the findings of the main analyses, with no significant associations across analyses (Supplementary Table 12). Similarly, the sensitivity analyses using recurrent depression as an outcome replicated the main analyses. Only

Table 1 | Results of Univariable Mendelian Randomization Analyses with Continuous Estradiol Levels

Exposure	Exposure GWAS and Data Source	Outcome	Outcome GWAS and Data Source	Threshold*	n SNPs	b	se	p	p _{FDR}
Combined pre- and postmenopausal sample from the UKB	Newly run; UKB excluding brain age gap sample	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-6}$	16	-0.18	0.28	.54	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-6}$	15	0.43	0.23	.06	.78
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-6}$	16	-0.03	0.07	.69	.98
Postmenopausal LIFE samples	Pott et al. ⁴⁵ ; LIFE-Adult & LIFE-Heart	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-6}$	10	0.11	0.17	.50	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-6}$	10	-0.04	0.13	.78	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-6}$	10	0.05	0.04	.30	.98
Premenopausal sample from the UKB	Haas et al. ²⁴ ; UKB	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-6}$	13	-0.42	0.31	.18	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-6}$	13	0.16	0.25	.51	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-6}$	13	-0.01	0.07	.93	.98
Postmenopausal sample from the UKB	Haas et al. ²⁴ ; UKB	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-6}$	12	0.08	0.11	.46	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-6}$	12	0.06	0.11	.62	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-6}$	12	-0.03	0.02	.20	.98

Note. Analyses were conducted using the inverse-variance weighted method (one-sided test). *p*_{FDR}, *p*-values adjusted across the 91 tests using FDR correction⁶¹. **p*-value threshold used as a cut-off to select instrumental variables (see Supplementary Note 3 for details). GWAS Genome-wide association study. UKB UK Biobank. MoBa Norwegian Mother and Child Cohort Study. PGC Psychiatric Genomics Consortium. ADGC Alzheimer's Disease Genetics Consortium. n SNPs Number of instrumental variables included in the respective analysis.

Table 2 | Results of Univariable Mendelian Randomization Analyses with Factors Related to Lifetime Estradiol Exposure

Exposure	Exposure GWAS and Data Source	Outcome	Outcome GWAS and Data Source	Threshold*	n SNPs	b	se	p	p _{FDR}
Reproductive Span	Schindler et al. ⁹ ; UKB excluding brain age gap sample	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-8}$	73	-0.02	0.13	.88	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-8}$	68	-0.04	0.09	.62	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-8}$	73	0.03	0.03	.27	.98
Age at Menarche	Perry et al. ⁵⁰ ; ReproGen	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-8}$	54	-0.30	0.15	.05	.76
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-8}$	54	-0.08	0.04	.04	.76
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-8}$	204	-0.05	0.06	.41	.98
Age at Menopause	Day et al. ⁴⁹ ; ReproGen	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-8}$	37	-0.00	0.04	.92	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-8}$	38	-0.00	0.01	.92	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-8}$	68	-0.01	0.02	.62	.98
Number of Childbirths	Elsworth et al. ⁵² ; UKB	Brain age gap	Newly run; UKB	$p < 5 \times 10^{-8}$	8	1.11	1.47	.45	.98
		Alzheimer's disease	Wang et al. ⁴⁷ ; ADGC	$p < 5 \times 10^{-8}$	8	-0.24	0.50	.63	.98
		Depression	Oppenheimer et al. ⁵¹ ; UKB, MoBa, PGC & iPSYCH	$p < 5 \times 10^{-8}$	8	0.25	0.25	.32	.98

Note. Analyses were conducted using the inverse-variance weighted method (one-sided test). *p*_{FDR}, *p*-values adjusted across the 91 tests using FDR correction⁶¹. **p*-value threshold used as a cut-off to select instrumental variables (see Supplementary Note 3 for details). GWAS Genome-wide association study. UKB UK Biobank. MoBa Norwegian Mother and Child Cohort Study. PGC Psychiatric Genomics Consortium. ADGC Alzheimer's Disease Genetics Consortium. ReProGen Reproductive Genetics Consortium. n SNPs Number of instrumental variables included in the respective analysis.

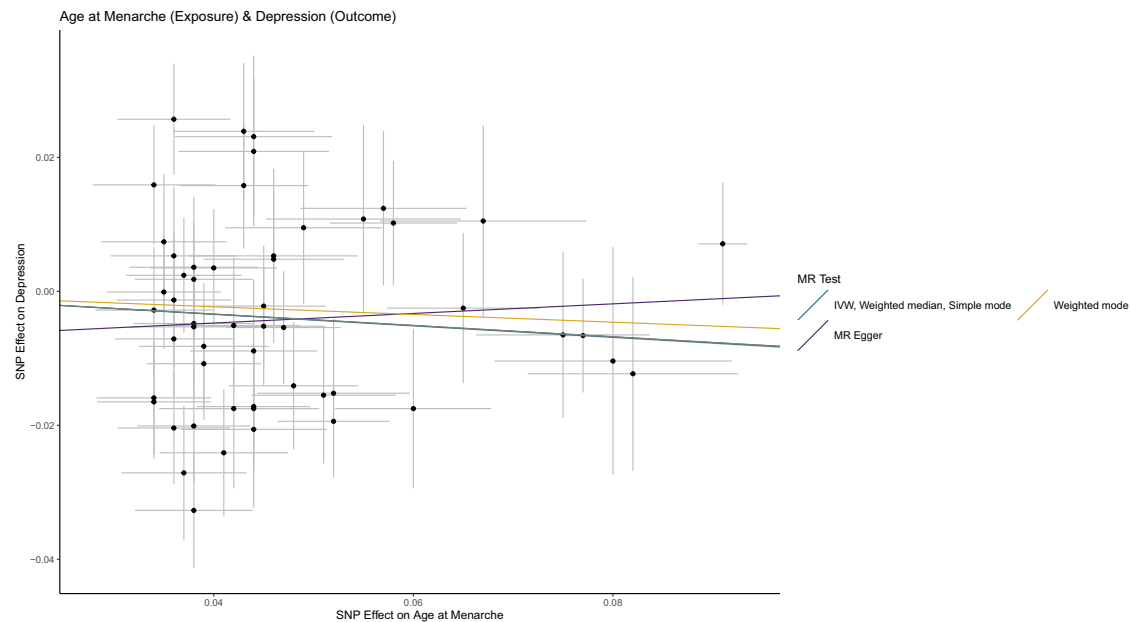


Fig. 3 | Effect of instrumental variables on age at menarche (exposure) and depression (outcome). Each point represents one single nucleotide polymorphism (SNP) included as an instrumental variable. Data are presented as the beta value and the confidence interval (\pm standard error) of the effect of a SNP on the exposure on the x-axis and the effect of a SNP on the outcome on the y-axis. The inverse-variance weighted (IVW), weighted median, and simple mode methods are plotted together for visualisation purposes, due to their overlapping regression

lines (see Table 2 and Supplementary Table 9 for details). The IVW estimate was used as the main estimation method (one-sided test) and was significant only before adjusting for multiple comparisons. Age at menarche⁵⁰: $N = 182,416$ females. Depression⁵¹: $N = 329,476$ females (44,610 cases with diagnosed depression and 284,866 healthy controls). See Table 3 for details on the included samples. *MR* Mendelian randomization. Source data are provided as a Source Data file.

the association with age at menarche as an exposure was significant, however, this association did not remain significant after adjusting for multiple comparisons ($b = -0.14$, $se = 0.05$, $p = .01$, $p_{FDR} = .30$) and was not robust across the estimation methods.

There was significant heterogeneity in the analyses of age at menarche as an exposure and depression as an outcome ($p < .001$), number of childbirths as an exposure and depression ($p = .02$) and brain age gap ($p < .001$) as outcomes, continuous estradiol levels measured in the postmenopausal sample from the UKB as an exposure and Alzheimer's disease as an outcome ($p = .02$) and reproductive span ($p = .04$) with brain age gap as an outcome, pointing to potential pleiotropic effects. Further, there was significant heterogeneity in the analyses of binary estradiol levels in the combined pre- and postmenopausal sample ($p = .04$), reproductive span ($p < .01$), and number of childbirths ($p = .02$) as exposures with recurrent depression as an outcome and estradiol levels in males with depression ($p < .01$; Supplementary Table 14).

In multivariable analyses, genetically-predicted BMI and continuous estradiol levels in the combined pre- and postmenopausal sample were not significantly associated with brain age gap or Alzheimer's disease as outcomes (Supplementary Table 13). Results were consistent when using binary estradiol levels as an exposure and when using multivariable MR-Egger. For depression as an outcome, continuous estradiol levels in the combined pre- and postmenopausal sample were not significant, but BMI was significant ($b = 0.14$, $se = 0.05$, $p = .01$, $p_{FDR} = .30$) in multivariable analyses. This was not robust when using multivariable MR-Egger, however, the result was consistent when using binary estradiol levels as an exposure (BMI: $b = 0.15$, $se = 0.05$, $p = .003$, $p_{FDR} = .27$). Furthermore, when including BMI, the association between age at menarche and depression, as well as the association between age at menarche and recurrent depression were not significant. This remained consistent when using multivariable MR-Egger. No significant heterogeneity was found in any of the analyses, except for the analysis using age at

menarche and BMI as exposures and depression as an outcome (Supplementary Table 14).

Discussion

We took a comprehensive approach to investigating potential causal associations between genetically predicted estradiol levels as well as factors related to lifetime estradiol exposure and genetically predicted brain age gap, Alzheimer's disease, and depression in females, by using Mendelian randomization. To this end, we conducted sex-stratified GWAS of brain age gap and female-specific GWAS of estradiol levels. The two-sample Mendelian randomization analyses revealed no significant causal associations between estradiol levels, reproductive span, age at menopause, and number of childbirths as exposures with brain age gap, Alzheimer's disease, and depression as outcomes. The results on estradiol levels were consistent across samples when using binary estradiol levels and across robust estimation methods. Using age at menarche as an exposure, a significant causal association with depression as an outcome was found, linking a younger age at menarche with a higher risk for depression. However, this result was not robust when adjusting for multiple comparisons, across estimation methods, or when including BMI in multivariable Mendelian randomization analyses.

In this study, we ran and annotated a sex-stratified GWAS on brain age gap, using grey matter and white matter MRI features from the UKB. We identified 2 significant loci on chromosomes 5 and 17 in the female-only sample. The significant locus on chromosome 17 has been reported by previous studies that conducted GWAS on brain age gap in mixed-sex samples^{26–30}. This locus includes the *MAPT* gene, which encodes the tau protein involved in several neurodegenerative diseases, including Alzheimer's disease³¹. The significant locus on chromosome 5 has not been reported in previous GWAS on brain age gap in mixed-sex samples^{26–30}. This region includes the *TENM2* gene, which has been linked to various traits, including smoking initiation, BMI, educational attainment, depression, age at menarche, and Alzheimer's disease in females (see

NHGRI-EBI GWAS Catalogue³²). Thus, this locus might be female-specific in its association with brain age gap. The absence of significant loci in the male-only sample suggests potential statistical power issues. Our findings emphasise the importance of conducting further, large-scale studies investigating sex-specific effects or gene-sex interactions in the genetic determinants of brain age gap.

Further, we ran and annotated a GWAS on estradiol levels in females from the UKB to avoid sample overlap. We identified two significant genomic loci, of which one (on chromosome 17) has been previously reported by a study on estradiol levels in females in the UKB²⁴. This locus is near the sex-hormone binding globulin (*SHBG*) gene, which is closely linked to concentrations of SHBG, a glycoprotein that binds to estradiol and testosterone^{23,24,32}. The independent significant SNPs identified in our GWAS have been linked to various related traits, including SHBG levels and testosterone levels (see Supplementary Table 15). In addition, the present study identified a significant locus on chromosome 19 that has been associated with white blood cell count and height, but not estradiol levels (see NHGRI-EBI GWAS Catalogue³²). Differences between the present study and previous studies may partly be traced back to differences in covariates and sample selection. The GWAS of the present study might be limited by the included covariates. For instance, while we controlled for menopausal status as well as history of oral contraceptive use, HRT use, hysterectomy, and oophorectomy, menstrual cycle phase was not considered due to the inclusion of pre- and postmenopausal females, potentially influencing the findings.

As a sensitivity analysis, we conducted a GWAS on binary estradiol levels (below or above detection limit), which had a substantially larger sample size. Similar to Haas et al.²⁴, we identified different significant loci in the binary estradiol GWAS, suggesting that different traits might be measured using these different approaches. Issues concerning estradiol measurements in the UKB have been previously discussed, including a potential bias towards the detection of loci associated with menopause due to the age of the participants and the substantial number of measurements below the detection limit^{23,24,33}. This may influence the findings of the present study, as the detected loci of the GWAS, as well as the selected instruments for estradiol levels, may be associated with related traits. This raises the possibility of confounding from horizontal pleiotropy¹⁵, where the genetic instruments influence the outcome through pathways other than the exposure (estradiol), such as via menopausal status or SHBG levels. Many large-scale databases, including the UKB, are conducted on aging cohorts and are confounded by survivor and healthy volunteer biases³⁴. Therefore, we conducted sensitivity analyses across different samples and approaches. Nevertheless, the present study is limited by the available data on estradiol levels and further highlights the need for large-scale, precise measurements conducted in diverse samples and age groups under consideration of female-specific variables – a data gap which has been repeatedly identified³⁵.

We did not find support for our hypotheses, which were largely based on findings from observational studies^{5–7,9}. Some similar patterns have been found in previous Mendelian randomization studies, for example, showing no causal associations between estradiol levels and depression²⁰, age at menopause and depression²², and age at menopause and age at menarche and risk for Alzheimer's disease¹⁸. Taken together with the results of these previous Mendelian randomization studies, our findings suggest that estradiol exposure may not directly influence brain and mental health, thus highlighting the need for careful consideration of possible confounding factors and reverse causation when interpreting observational studies. However, it must be mentioned that some of the additionally analysed exposure variables, such as HRT use and oral contraceptive use, might not have a substantial genetic component, potentially leading to null findings in the present study. Furthermore, non-linear and age-dependent effects, as have been suggested for associations of the number of childbirths

and HRT use with brain health^{7,36}, should be investigated. For instance, previous studies highlight the importance of timing, duration, and formulation of HRT use³⁶. These aspects were not captured by the variable used for HRT use in the present study, possibly resulting in null findings. Further, the GWAS of the history of hysterectomy did not take important influencing factors into account, such as the medical reason for the procedure, the timing of the procedure (i.e., before or after menopause), and whether an oophorectomy also took place. These limitations can be traced back to the use of GWAS from consortia, which have advanced research through publicly available, large-scale datasets, however, are often designed in a standardised matter across sexes and may not consider potentially relevant covariates for specific variables or samples. Future research is needed to disentangle genetic and environmental contributions to these variables and further assess their causal associations with brain and mental health.

Moreover, the generalisability of the present findings is limited due to the restriction of the analyses to White European samples, which is a result of the limited availability of diverse large-scale datasets. Females from different ethnic groups have been suggested to differ in their hormonal profiles³⁷, thus the present findings should be replicated across diverse samples. Furthermore, it is important to note that while we set a strict threshold for selecting instrumental variables in the Mendelian randomization analyses to satisfy the relevance assumption, the exchangeability and exclusion restriction assumptions cannot be verified and may limit the validity of the results¹⁵. Although we conducted a variety of sensitivity analyses and reduced sample overlap, pleiotropic effects and confounding of the genetic variant-outcome association remain possible, for example, arising from population stratification or assortative mating¹⁵. This is especially important to note for the analyses that exhibited significant heterogeneity, such as the analysis with age at menarche as an exposure and depression as an outcome, or weak instrument bias, such as estradiol levels in the multivariable analyses including BMI as an additional exposure. Further, power issues might be present in the analyses with few instrumental variables, especially in the sensitivity analyses using methods with lower power, such as MR-Egger, and in the presence of weak instrument bias³⁸.

Previous Mendelian randomization studies have reported a causal link between a younger age at menarche and an increased risk for depression in adolescents and adults^{16,17,19,21,22}. In the multivariable Mendelian randomization analyses of the present study, the attenuation of the significant effect when including BMI suggests potential pleiotropy, whereby the genetic variants included in the analysis with age at menarche influence depression through BMI. BMI is known to be linked to an earlier onset of puberty and, in line with the present study, has previously been highlighted as a relevant confounder in the causal relationship between a younger age at menarche and depression^{16,19}. Previous studies have also discussed possible age-dependent effects of age at menarche on depression^{16,19}, warranting exploration to explain inconsistent findings.

The female lifespan is marked by fluctuations in estradiol levels throughout the menstrual cycle, the peripartum period, and the perimenopausal period, and these fluctuations have repeatedly been linked to brain and mental health, largely in observational studies³. For instance, brain age gap has been found to vary across the menstrual cycle, with higher estradiol levels linked to a younger brain age gap³⁹ and estradiol fluctuations have been linked with changes in brain regions that are implicated in depression and Alzheimer's disease⁴⁰. The emergence of female-specific subtypes of depression, including premenstrual dysphoric disorder, peripartum depression, and perimenopausal depression, follow a similar trajectory to hormonal changes³. For instance, increased brain plasticity during the peripartum period has been linked with changes in estradiol levels and an increased vulnerability to mental disorders⁴¹. Furthermore, progesterone, as another prevalent sex steroid in females, has been found to

have varying effects on mood, with suggested protective effects in the peripartum period and possible symptom-provoking effects in premenstrual dysphoric disorder⁴². Future studies should use the present framework to further explore the role of other sex steroids on brain and mental health in females. Together, these studies suggest that sex steroids differently modulate females' brain and mental health across the lifespan, and that hormonal fluctuations rather than absolute hormone levels might be pivotal in this dynamic.

We did not find significant effects of estradiol levels on brain and mental health in males, who are known to experience substantially lower estradiol fluctuations, further supporting the absence of a constant effect of estradiol. If the effect of estradiol on brain health and mental health is phasic and not constant, this might have influenced the current results and contributed to the non-significant findings. Due to the methods used, the present study was unable to examine the potential effects of estradiol fluctuations or susceptible periods to estradiol exposure throughout the female lifespan. Therefore, we used the phrase lifetime estradiol exposure. However, the sample used for our main analysis on estradiol levels consisted predominantly of postmenopausal females, thus likely not capturing effects across the full lifespan. Ideally, to capture time-varying effects, multivariable Mendelian randomization should be applied using GWAS summary statistics across multiple time points⁴³. For instance, estradiol levels should be assessed repeatedly across the female lifespan. However, this approach is difficult to implement due to the limited availability of the required data. Nevertheless, a previous study has shown age-specific effects of genetic associations of age at menopause and revealed different magnitudes of causal effects dependent on the age group⁴⁴, emphasising that the genetic underpinnings of female reproductive factors vary across the lifespan. Thus, future studies are needed that assess estradiol levels in females across multiple time-points – across the menstrual cycle, across pregnancy, or across the lifespan – and that examine these effects using combined sex-specific and causal approaches.

While the non-significant associations between estradiol levels and brain age gap, depression, and Alzheimer's disease found in the present study suggest that estradiol might not have a simple direct influence on brain health and mental health in females, associations are complex and require investigation using a more parcellated lifespan approach. Future research should consider the causal effects of hormonal fluctuations and periods of increased susceptibility to sex steroids. Further, sex-specific genetic data is needed to disentangle the complex relationships underlying sex differences in brain health and mental health and to ultimately aid the advancement of individualised healthcare.

Methods

Samples and data sources

Available summary statistics. Summary statistics from GWAS, including only White European female samples, were used for all variables, except for sensitivity analyses conducted in White European males. Samples were selected to minimise overlap between the exposure and outcome samples, while allowing for the largest possible sample sizes. No statistical method was used to predetermine sample size. Main exposure variables of interest included estradiol levels and factors related to lifetime estradiol exposure (i.e., reproductive span, age at menarche, age at menopause, and number of childbirths). The main outcome variables of interest were brain age gap, Alzheimer's disease, and depression.

To ensure robustness to effects of menopausal status previously shown to impact GWAS results²⁴, several datasets were included for estradiol levels. The following samples from the UKB were included: a combined sample of pre- and postmenopausal females, a sample of only premenopausal females, and a sample of only postmenopausal females²⁴. Main analyses in each sample were performed using GWAS of continuous estradiol levels. Sensitivity analyses were conducted in

each sample using a binary approach, examining estradiol levels above and below the detection limit²⁴ to increase sample sizes. Further, analyses on continuous estradiol levels were replicated in an independent sample of postmenopausal females from the LIFE-Adult and LIFE-Heart studies⁴⁵ and a male-only sample from the UKB⁴⁶.

For Alzheimer's disease⁴⁷ as an outcome, a larger GWAS for age at menopause was used that was not limited to natural menopause⁴⁸ and a sensitivity analysis was conducted using age at natural menopause⁴⁹. To avoid sample overlap, analyses using age at menarche⁵⁰ and age at natural menopause⁴⁹ as exposures and brain age gap and depression⁵¹ as outcomes were conducted using GWAS summary statistics from a smaller, independent sample. Supplementary analyses were conducted using factors related to exogenous hormone use and health-related procedures likely impacting estradiol levels, including oral contraceptive use, HRT use, history of hysterectomy, and history of oophorectomy⁵² as exposure variables. As the GWAS on depression⁵¹ included both single episode and recurrent depression, sensitivity analyses were conducted using recurrent depression⁵³ as an outcome variable to check whether disease burden influenced the results found. Further, as some of the analyses using depression⁵¹ as an outcome had sample overlap, these analyses were replicated using a subsample of the depression GWAS that did not include the UKB sample⁵³. BMI⁵⁴ was included in multivariable sensitivity analyses, to account for potential pleiotropic effects^{16,17,21}.

For estradiol levels in the independent replication sample⁴⁵, estradiol levels in the male-only sample⁴⁶, age at menarche^{48,50}, age at (natural) menopause^{48,49}, number of childbirths, oral contraceptive use, HRT use, history of hysterectomy, history of oophorectomy⁵², and BMI⁵⁴ summary statistics were publicly available in data repositories or obtained from the IEU OpenGWAS project⁵². For estradiol levels in the premenopausal sample and in the postmenopausal sample²⁴, reproductive span⁹, Alzheimer's disease⁴⁷ in the female-only and male-only samples, depression⁵¹, the depression subsample excluding UKB⁵³ in the female-only and male-only samples, and recurrent depression⁵³ summary statistics were requested and received from the respective authors. For depression in the male-only sample, a GWAS meta-analysis was run in line with the female-only GWAS⁵¹ including data from the UKB¹⁰, MoBa²⁵, and a previously conducted GWAS, received from the authors upon request⁵³ (see Supplementary Notes 1, 2, and 5 for details). See Table 3 and Supplementary Table 1 for a complete overview of the samples and datasets included in the present study.

UKB samples. As, to the best of our knowledge, no previously conducted sex-stratified GWAS on brain age gap existed, we ran separate GWAS on female- and male-only samples using data from the UKB. Further, to avoid sample overlap in the analyses using estradiol levels in the combined sample of pre- and postmenopausal females, we ran a female-specific GWAS using data from the UKB. The UKB complies with the Helsinki Declaration, with informed consent obtained from all participants. For the male-only GWAS on brain age gap, only genetic males (XY chromosome) with White British ancestry were included. For the female-only GWAS on brain age gap and estradiol levels, only genetic females (XX chromosome) with White British ancestry were included. For both GWAS on brain age gap, only participants with both T1-weighted and diffusion-weighted MRI data were included, and participants with ICD-10 diagnoses known to impact brain health were excluded (see Supplementary Note 1 for details on exclusions). For the GWAS on estradiol levels, participants with serum blood sample data from their first visit to the UKB assessment centre (2006 – 2010) were included. To avoid sample overlap, participants were excluded from the estradiol levels sample if they were included in the brain age gap sample. Further, participants who answered, "prefer not to answer" or "do not know/not sure" on any of the covariates included in the GWAS on continuous estradiol levels were excluded (Supplementary Note 1).

Table 3 | Included Datasets and Samples

Variable	GWAS	Sample Description	Data Source	Definition of Phenotype
Estradiol Levels (continuous approach; combined pre- and postmenopausal sample)	Newly conducted	N = 34,697	UKB	Estradiol levels above detection limit (range: 175 – 14,588 pmol/L; mean = 534.11 pmol/L)
Estradiol Levels (continuous approach; premenopausal sample)	Haas et al. ²⁴ – received upon request	N = 33,032	UKB	Estradiol levels above detection limit
Estradiol Levels (continuous approach; postmenopausal sample)	Haas et al. ²⁴ – received upon request	N = 3759	UKB	Estradiol levels above detection limit
Estradiol Levels (continuous approach; postmenopausal replication sample)	Pott et al. ⁴⁵ – publicly available	N = 2607	LIFE-Adult & LIFE-Heart	Estradiol levels in postmenopausal females
Brain Age Gap	Newly conducted	N = 14,287	UKB	The difference between predicted brain age and chronological age
Alzheimer's Disease	Wang et al. ⁴⁷ – received upon request	N = 12,772 (5,705 cases and 7,067 controls)	ADGC phase 1 & 2	Diagnosis of Alzheimer's disease
Depression	Oppenheimer et al. ⁵¹ – previously conducted	N = 329,476(44,610 cases & 284,866 controls)	UKB, MoBa, & PGC & iPSYCH	DSM-IV or ICD-10 (F32 or F33) diagnosis of major depressive disorder
Reproductive Span	Schindler et al. ⁹ – received upon request	N = 121,620	UKB	Age at natural menopause minus age at menarche
Age at Menarche	Loh et al. ⁴⁸ – publicly available	N = 279,470	UKB	Age when periods started (menarche)
Age at Menarche*	Perry et al. ⁵⁰ – publicly available	N = 182,416	ReproGen	Age at menarche
Age at Menopause	Loh et al. ⁴⁸ – publicly available	N = 156,364	UKB	Age at menopause (last menstrual period)
Age at Natural Menopause*	Day et al. ⁴⁹ – publicly available	N = 69,360	ReproGen	Age at natural (non-surgical) menopause
Number of Childbirths	Elsworth et al. ⁵⁷ – publicly available	N = 250,782	UKB	Number of live childbirths

Note. All GWAS used for the main analyses were conducted in White European female samples. *To avoid sample overlap, different datasets with smaller sample sizes were used in the analyses with brain age gap and depression as outcomes. See Supplementary Table 1 for datasets used for sensitivity and supplementary analyses. UKB UK Biobank. ADGC Alzheimer's Disease Genetics Consortium. MoBa The Norwegian Mother and Child Cohort Study. PGC Psychiatric Genomics Consortium. ReproGen Reproductive Genetics Consortium.

Sample overlap. Sample overlap was avoided for all analyses using Alzheimer's disease and recurrent depression as outcomes, as well as the sensitivity analyses in the depression⁵³ subsample, excluding UKB. Sample overlap with brain age gap and depression as outcomes was avoided for age at menarche and age at menopause as exposures by using datasets that did not include the UKB sample. Further, sample overlap with brain age gap as an outcome was avoided for estradiol levels (combined pre- and postmenopausal samples) and reproductive span as exposures by excluding the MRI sample from the UKB. However, sample overlap could not be avoided for the analyses using continuous estradiol levels in the combined pre- and postmenopausal sample and reproductive span as exposures and depression as an outcome (maximum sample overlap: 10.53% for estradiol levels and 36.91% for reproductive span). Further, there was sample overlap for continuous estradiol levels in the premenopausal sample and in the postmenopausal sample with brain age gap and depression (maximum sample overlap: 43.25% for premenopausal females with brain age gap as an outcome; 10.03% for premenopausal females with depression as an outcome; 26.31% for postmenopausal females with brain age gap as an outcome; 1.14% for postmenopausal females with depression as an outcome). For the number of childbirths, there was a maximum sample overlap of 5.70% with brain age gap as an outcome and 82.19% with depression as an outcome. Further, sample overlap could not be avoided for the supplementary analyses using factors related to exogenous hormone use and health-related procedures and the sensitivity analyses using estradiol levels in males with brain age gap and depression as outcomes, as well as using binary estradiol levels with depression as an outcome (see Supplementary Table 2).

Brain age gap prediction

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.⁵⁵ and Miller et al.⁵⁶ for details on MRI data acquisition and protocols in the UKB). Automated surface-based morphometry and subcortical segmentation pipelines in FreeSurfer v5.3⁵⁷ were used to process raw T1-weighted images. The standard set of FreeSurfer-derived subcortical and cortical summary statistics⁵⁷ were further supplemented by a fine-grained cortical parcellation scheme⁵⁸ to extract cortical thickness, area, and volume for 180 regions of interest per hemisphere. Outliers based on Euler numbers more than 4 *SD* above the mean were excluded, to avoid poor-quality data likely arising from movement⁵⁹. The data from FreeSurfer was controlled for scanner site and intracranial volumes using linear models. Diffusion-weighted data were processed using an optimised diffusion pipeline^{60–62} and similarly controlled for scanner site using linear models. The diffusion-weighted data passed TBSS post-processing quality control through the YTRIUM algorithm^{60–62}. In total, 1,118 T1-weighted features and 912 diffusion-weighted features were included in the brain age prediction model.

Python (v3.7.4) was used for the brain age prediction. In line with previous studies on brain age gap in females^{7–9}, brain age was computed using the *XGBoost* (eXtreme Gradient Boosting) regression model, based on a decision-tree ensemble algorithm⁶³. Hyperparameters were tuned using a nested cross-validation with 5 inner folds for randomised search and 10 outer folds for validation of the model (see **general model setup**; see Supplementary Table 3 for model performance metrics). Brain age was estimated for each individual, and brain age gap was calculated by subtracting chronological age from the brain age prediction¹¹. A final sample of *N* = 14,287 females was included in the brain age gap calculations and the subsequently conducted GWAS (age range: 45.13 – 81.83 years; mean = 63.56, *SD* = 7.27; see Supplementary Note 1 for details on the male brain age gap sample).

Estradiol levels

In the UKB, estradiol levels were measured using two-step competitive analysis on a *Beckman Coulter Unicel Dxl 800* with a minimum detection limit of 175 pmol/L (for details see [UKB Data Field 30800](#)). A majority of the participants in the UKB have estradiol levels below this limit, likely due to the advanced age of the cohort. The GWAS on estradiol levels using the continuous approach included females with measurements above the detection limit ($N = 34,697$; age range: 40.16 – 70.17 years; mean = 48.30, $SD = 5.96$). As the estradiol levels were positively skewed, we performed an inverse rank normalisation on the data. Following previous studies^{23,24}, a second GWAS was conducted using a binary approach which classified all females ($N = 207,119$) as either below (0) or above (1) the detection limit to increase the sample size (Supplementary Note 1).

GWAS procedure

The UKB v3 imputed genetic data was used, which has been genotyped, extensively quality controlled, and imputed by the UKB genetics team (see Bycroft et al.⁶⁴ for details). PLINK 2.0⁶⁵ was used to run the GWAS. We performed standard quality check procedures, by excluding individuals with more than 10% missingness and setting the minor allele frequency (MAF) threshold at .005. We filtered out single nucleotide polymorphisms (SNPs) with more than 5% missingness as well as those out of the Hardy-Weinberg equilibrium (HWE) at $p < 1 \times 10^{-9}$. GWAS were run using the standard additive model of linear associations for the continuous variables (i.e., brain age gap and continuous estradiol levels) and logistic associations for the binary variable (i.e., binary estradiol levels). All analyses were controlled for the age of the participants at the time of measurement and the first 20 principal components to account for population structure. The GWAS on continuous estradiol levels was further controlled for menopausal status (premenopausal/postmenopausal), history of oral contraceptive use (yes/no), history of HRT use (yes/no), history of bilateral oophorectomy (yes/no), and history of hysterectomy (yes/no), as these variables are known to influence estradiol levels.

Post-GWAS annotations. For annotations of the GWAS results, Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) was used, applying the *SNP2GENE* function⁶⁶. GWAS results were clumped using the 1000 Genomes projects phase 3 European dataset (1KG/Phase3) linkage disequilibrium (LD) structure with a maximum p -value of lead SNPs $< 5 \times 10^{-8}$, a maximum p -value cutoff < 0.05 , an R^2 threshold $\geq .6$ to define independent significant SNPs, an R^2 threshold $\geq .1$ to define lead SNPs, and 250 kilo-bases (kb) as the maximum distance to merge LD blocks into a locus⁶⁶. Positional mapping was performed using a maximum distance of 10 kb. Independent significant SNPs for each genomic region were identified, along with lead SNPs defined as SNPs with the smallest p -value for a genomic region. For estradiol levels, independent significant SNPs were linked to traits as reported in the GWAS Catalogue³² to facilitate investigation of potential pleiotropy of these variables used as exposures in the Mendelian randomization analyses. Using the *GENE2FUNC* function implemented by FUMA, the expression levels of the mapped genes were investigated for each of the GTEx v8 54 tissue types⁶⁷ and hypergeometric tests were performed for gene-set enrichment analyses, testing whether genes are overrepresented in any of the pre-defined gene sets and their associations with various phenotypes⁶⁶.

Two-sample Mendelian randomization analyses

Two-sample Mendelian randomization analyses were performed using R (v4.3.1). The main steps of the analyses are visualised in Fig. 4. All p -values for the IVW estimates were adjusted for multiple comparisons across all analyses (91 comparisons), including the supplementary, sensitivity, and follow-up analyses, using false discovery rate (FDR; 5%) correction⁶⁸. The present study was preregistered on Open Science

Framework (OSF; see Supplementary Note 8 for deviations from the preregistration).

Selection of instrumental variables. The instrumental variables included in a Mendelian randomization analysis are assumed to be associated with the respective exposure variable (relevance assumption), not be associated with the outcome variable via confounding pathways (exchangeability assumption), and not affect the outcome variable in any way other than through the exposure variable (exclusion restriction assumption)¹⁵. The relevance assumption can be verified, while the exchangeability and exclusion restriction assumptions are not verifiable¹⁵, however, a set of sensitivity analyses, including the use of robust methods and multivariable Mendelian randomization, were conducted to assess biases arising from potential violations.

To satisfy the relevance assumption, instrumental variables were selected by applying a genome-wide significance threshold of $p < 5 \times 10^{-8}$ ($p < 5 \times 10^{-7}$ or $p < 5 \times 10^{-6}$ in the case of few available SNPs; see Supplementary Note 3 for details on the selection of instrumental variables for each analysis) to identify SNPs associated with the respective exposure variables. Using the *TwoSampleMR* package (v0.5.11)⁶⁹, which makes use of data from the 1000 Genomes project, significant SNPs in LD were pruned to include only the most significant SNP, and the exposure and outcome datasets were harmonised.

First-stage F-statistics were approximated for all instrumental variables as a measure of instrument strength ($F < 10$ indicating potential weak instrument bias), using the *MendelianRandomization* package (v0.9.0)⁷⁰. The instrumental variables did not indicate weak instrument bias for any of the main analyses. For the sensitivity analyses, weak instrument bias was only indicated for the instruments used for the number of childbirths as an exposure with the depression⁵³ subsample excluding UKB as an outcome ($F = 0.80$) and with recurrent depression as an outcome ($F = 0.70$; Supplementary Table 4). Furthermore, Cochran's Q was computed as a measure of heterogeneity, which assesses whether the instrumental variables underlie the same causal parameter and can thereby be used as an indicator of potential pleiotropy (i.e., instrumental variables affect the outcome through other biological pathways) when significant¹⁵.

Univariable analyses. Univariable two-sample Mendelian randomization analyses were conducted using the GWAS summary statistics. Univariable analyses were performed between all separate exposure variables and all separate outcome variables (Fig. 4). Odds ratios were log-transformed before conducting analyses. Analyses were performed using the *TwoSampleMR* package (v0.5.11)⁶⁹. For the main analyses, the IVW method was applied, as it maximises statistical power¹⁵. Results were compared to more robust Mendelian randomization approaches, including MR-Egger, which allows for directional pleiotropy, as well as the weighted median, simple mode, and weighted mode estimation methods, which are more robust to outliers and make varying assumptions regarding the validity of the instrumental variables (see Burgess and Thompson, 2021¹⁵ for details). Using the *MRIap* package (v0.0.3.2)⁷¹, MRIap was performed for analyses with sample overlap. MRIap corrects for potential biases arising from sample overlap, weak instruments, and winner's curse (i.e., the overestimation of effects in discovery GWAS)⁷¹.

Multivariable analyses. Multivariable Mendelian randomization analyses were conducted, including continuous estradiol levels in the combined pre- and postmenopausal sample and BMI as exposure variables with brain age gap, Alzheimer's disease, and depression as outcomes. Sensitivity analyses were conducted using the binary estradiol approach and, to avoid sample overlap, using the depression subsample excluding UKB. Furthermore, based on our univariable findings and following previous studies^{16,17,21}, follow-up multivariable analyses were conducted, including age at menarche and BMI as

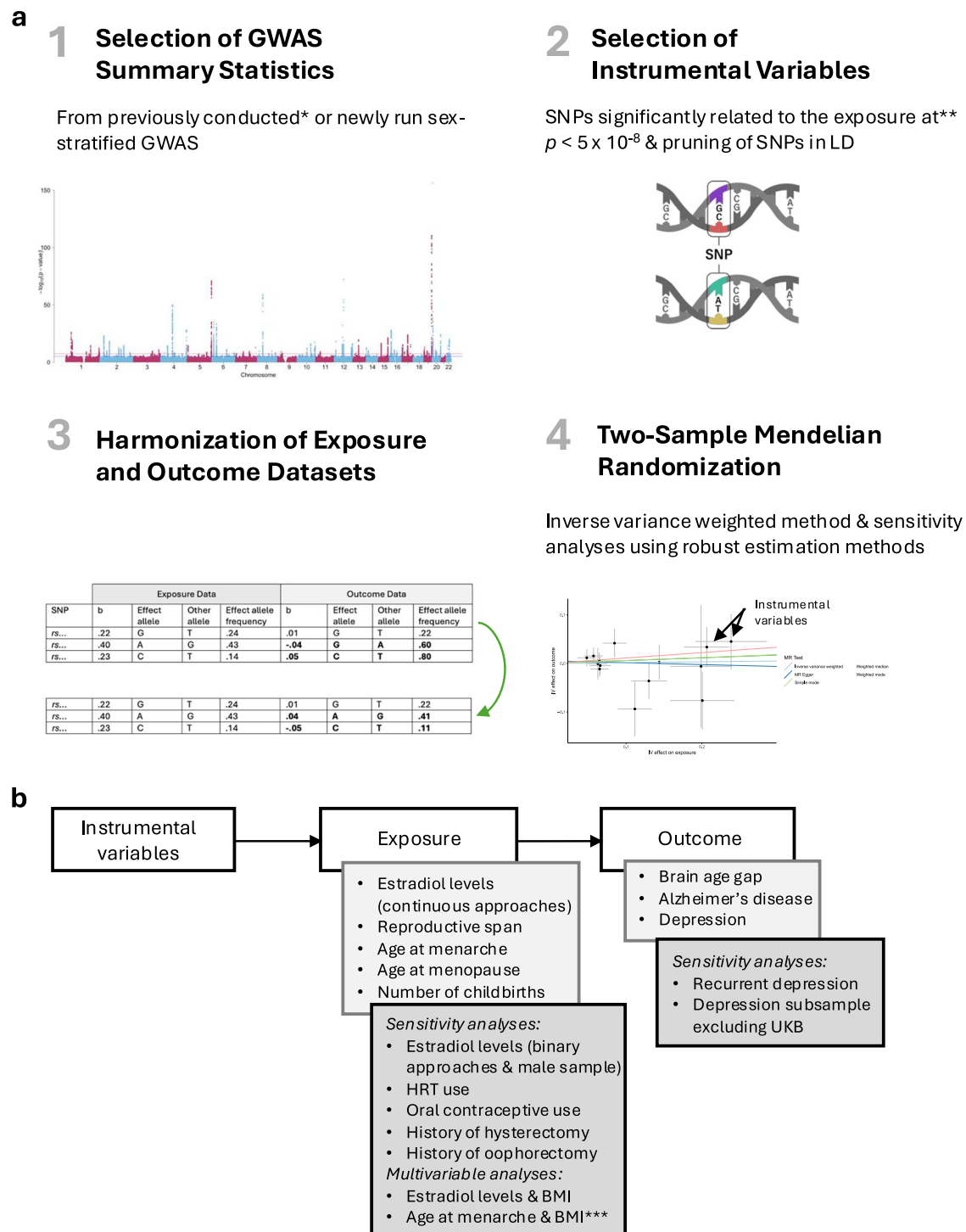


Fig. 4 | Analysis steps and conducted Mendelian randomization analyses.

a Analysis steps. 1: Example Manhattan plot of a genome-wide association study (GWAS) for visualisation purposes. The blue/lower and red/upper lines indicate the suggestive ($p < 5 \times 10^{-8}$) and the genome-wide significance ($p < 5 \times 10^{-8}$; i.e., accounting for multiple comparisons) thresholds, respectively. The instrumental variables included in a Mendelian randomization analysis are assumed to be associated with the respective exposure variable (relevance assumption), not be associated with the outcome variable via confounding pathways (exchangeability assumption), and not affect the outcome variable in any way other than through the exposure variable (exclusion restriction assumption)¹⁵. *Requested from respective authors or openly available. 2: To satisfy the relevance assumption, the genome-wide significance threshold (p -value cut-off) was chosen for the selection of instrumental variables, when possible. **Relaxed in the case of too few available instrumental variables (see Supplementary Note 3 and Supplementary Table 4 for

details). 3: Example data of exposure and outcome GWAS summary statistics for visualisation purposes. **b** refers to the beta-estimate of the GWAS for each single-nucleotide polymorphism (SNP). 4: Example plot of two-sample Mendelian randomization analyses for visualisation purposes. Each point represents one SNP included as an instrumental variable. Data are presented as the beta value and the confidence interval (+/- standard error) of the effect of a SNP on the exposure on the x-axis and the effect of a SNP on the outcome on the y-axis. Analyses were conducted using the inverse variance weighted method as a main estimation method (one-sided test). As the exchangeability and exclusion restriction assumptions are not verifiable¹⁵, sensitivity analyses were conducted to assess biases arising from potential violations. **b** Conducted analyses. ***Follow-up analysis with depression as an outcome. *LD* Linkage disequilibrium. *HRT* Hormone replacement therapy. *BMI* Body mass index. *UKB* UK Biobank.

exposures with depression, the depression subsample excluding UKB, and recurrent depression as outcomes. Using the *TwoSampleMR* package (v0.5.11)⁶⁹, instrumental variables were extracted from the exposure datasets, the exposures and the outcome were harmonised, and multivariable IVW analyses were run. Conditional F-statistics for instrument strength were computed, and the multivariable MR-Egger method was applied to check the robustness of the multivariable Mendelian randomization results using the *MVMR* package (v0.1)⁷². The instrumental variables for the multivariable analyses with estradiol levels and BMI as exposures indicated no weak instrument bias for BMI, however, possible weak instrument bias for estradiol levels (brain age gap and Alzheimer's disease as outcomes: conditional $F=1.92$ for the continuous approach and 8.87 for the binary approach in the combined pre- and postmenopausal sample from the UKB; depression as an outcome: conditional $F=1.93$ for the continuous approach and 9.16 for the binary approach in the combined pre- and postmenopausal sample from the UKB; depression subsample excluding UKB as an outcome: conditional $F=0.99$ for the continuous approach and 8.49 for the binary approach in the combined pre- and postmenopausal sample from the UKB; see also Supplementary Note 3 and Supplementary Table 4). The instrumental variables included in the multivariable analyses of age at menarche and BMI as exposures did not exhibit weak instrument bias, except for possible weak instrument bias for BMI in the analysis with the depression subsample excluding UKB as an outcome (conditional $F=9.97$).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

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 generated in this study have been deposited in the [NHGRI-EBI GWAS Catalog](#) under the following accession codes: 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 under restricted access from the [UK Biobank](#) and the [Norwegian Mother and Child Cohort Study](#). Access can be obtained from the respective databases upon request. 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 Table 3 and Supplementary Table 1. Source data are provided in this paper.

Code availability

The code used for the analyses is deposited on [GitHub](#) (<https://doi.org/10.5281/zenodo.17099458>)⁷³.

References

- Essau, C. A., Lewinsohn, P. M., Seeley, J. R. & Sasagawa, S. Gender differences in the developmental course of depression. *J. Affect. Disord.* **127**, 185–190 (2010).
- Ferretti, M. T. et al. Sex differences in Alzheimer disease — the gateway to precision medicine. *Nat. Rev. Neurol.* **14**, 457–469 (2018).
- Barth, C., Crestol, A., De Lange, A.-M. G. & Galea, L. A. M. Sex steroids and the female brain across the lifespan: insights into risk of depression and Alzheimer's disease. *Lancet Diabetes Endocrinol.* **12**, 926–941 (2023).
- Hara, Y., Waters, E. M., McEwen, B. S. & Morrison, J. H. Estrogen effects on cognitive and synaptic health over the lifecourse. *Physiol. Rev.* **95**, 785–807 (2015).
- Georgakis, M. K. et al. Association of age at menopause and duration of reproductive period with depression after menopause: a systematic review and meta-analysis. *JAMA Psychiatry* **73**, 139 (2016).
- Gong, J., Harris, K., Peters, S. A. E. & Woodward, M. Reproductive factors and the risk of incident dementia: A cohort study of UK Biobank participants. *PLoS Med.* **19**, e1003955 (2022).
- de Lange, A.-M. G. et al. Population-based neuroimaging reveals traces of childbirth in the maternal brain. *Proc. Natl. Acad. Sci. USA* **116**, 22341–22346 (2019).
- Crestol, A. et al. Linking menopause-related factors, history of depression, APOE $\epsilon 4$, and proxies of biological aging in the UK biobank cohort. *Hormones Behav.* **164**, 105596 (2024).
- Schindler, L. S. et al. Associations between abdominal adipose tissue, reproductive span, and brain characteristics in postmenopausal women. *NeuroImage: Clin.* **36**, 103239 (2022).
- Sudlow, C. et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci.* **40**, 681–690 (2017).
- Han, L. K. M. et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol. Psychiatry* **26**, 5124–5139 (2021).
- Kaufmann, T. et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623 (2019).
- Ambikairajah, A., Tabatabaei-Jafari, H., Hornberger, M. & Cherbuin, N. Age, menstruation history, and the brain. *Menopause* **28**, 167–174 (2021).
- Burgess, S. & Thompson, S. G. *Mendelian Randomization*. (Chapman and Hall/CRC, 2021).
- Askelund, A. D. et al. Assessing causal links between age at menarche and adolescent mental health: a Mendelian randomisation study. *BMC Med.* **22**, 155 (2024).
- Hirtz, R. et al. Causal effect of age at menarche on the risk for depression: results from a two-sample multivariable mendelian randomization study. *Front. Genet.* **13**, 918584 (2022).
- Li, M. et al. The role of age at menarche and age at menopause in Alzheimer's disease: evidence from a bidirectional mendelian randomization study. *Aging* **13**, 19722–19749 (2021).
- Sequeira, M.-E., Lewis, S. J., Bonilla, C., Smith, G. D. & Joinson, C. Association of timing of menarche with depressive symptoms and depression in adolescence: Mendelian randomisation study. *Br. J. Psychiatry* **210**, 39–46 (2017).
- Veeneman, R. R., Verweij, K. J. H., Sommer, I. E. C., Treur, J. L. & Vermeulen, J. M. The role of sex hormones in severe mental illness: a genetic exploration. Preprint at <https://doi.org/10.1101/2024.06.25.24309226> (2024).
- Wang, Z., Lu, J., Weng, W., Fu, J. & Zhang, J. Women's reproductive traits and major depressive disorder: A two-sample Mendelian randomization study. *J. Affect. Disord.* **326**, 139–146 (2023).
- Yu, Y. et al. Causal associations between female reproductive behaviors and psychiatric disorders: a lifecourse Mendelian randomization study. *BMC Psychiatry* **23**, 799 (2023).
- Schmitz, D. et al. Genome-wide association study of estradiol levels and the causal effect of estradiol on bone mineral density. *J. Clin. Endocrinol. Metab.* **106**, e4471–e4486 (2021).
- Haas, C. B., Hsu, L., Lampe, J. W., Wernli, K. J. & Lindström, S. Cross-ancestry genome-wide association studies of sex hormone concentrations in pre- and postmenopausal women. *Endocrinology* **163**, bqac020 (2022).
- Magnus, P. et al. Cohort profile update: the norwegian mother and child cohort study (MoBa). *Int. J. Epidemiol.* **45**, 382–388 (2016).

26. Jawinski, P. et al. Genome-wide analysis of brain age identifies 59 associated loci and unveils relationships with mental and physical health. *Nat. Aging* **5**, 2086–2103 (2025).
27. Jonsson, B. A. et al. Brain age prediction using deep learning uncovers associated sequence variants. *Nat. Commun.* **10**, 5409 (2019).
28. Leonardsen, E. H. et al. Genetic architecture of brain age and its causal relations with brain and mental disorders. *Mol. Psychiatry* **28**, 3111–3120 (2023).
29. Ning, K. et al. Improving brain age estimates with deep learning leads to identification of novel genetic factors associated with brain aging. *Neurobiol. Aging* **105**, 199–204 (2021).
30. Smith, S. M. et al. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *ELife* **9**, e52677 (2020).
31. Rademakers, R., Cruts, M. & Van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum. Mutat.* **24**, 277–295 (2004).
32. Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
33. Ruth, K. S. et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).
34. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
35. Heller, C. et al. The ENIGMA-Neuroendocrinology working group to bridge gaps in female mental health research. *Nat. Ment. Health* **2**, 348–350 (2024).
36. Barth, C. et al. Menopausal hormone therapy and the female brain: leveraging neuroimaging and prescription registry data from the UK Biobank cohort. *eLife* **13**, RP99538 (2025).
37. Kim, C. et al. Racial/ethnic differences in sex hormone levels among postmenopausal women in the diabetes prevention program. *J. Clin. Endocrinol. Metab.* **97**, 4051–4060 (2012).
38. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
39. Franke, K., Hagemann, G., Schleussner, E. & Gaser, C. Changes of individual BrainAGE during the course of the menstrual cycle. *NeuroImage* **115**, 1–6 (2015).
40. Dubol, M. et al. Neuroimaging the menstrual cycle: A multimodal systematic review. *Front. Neuroendocrinol.* **60**, 100878 (2021).
41. Barba-Müller, E., Craddock, S., Carmona, S. & Hoekzema, E. Brain plasticity in pregnancy and the postpartum period: links to maternal caregiving and mental health. *Arch. Women's Ment. Health* **22**, 289–299 (2019).
42. Sundström-Poromaa, I., Comasco, E., Sumner, R. & Luders, E. Progesterone – Friend or foe? *Front. Neuroendocrinol.* **59**, 100856 (2020).
43. Sanderson, E., Richardson, T. G., Morris, T. T., Tilling, K. & Davey Smith, G. Estimation of causal effects of a time-varying exposure at multiple time points through multivariable mendelian randomization. *PLoS Genet.* **18**, e1010290 (2022).
44. Ojavee, S. E. et al. Genetic insights into the age-specific biological mechanisms governing human ovarian aging. *Am. J. Hum. Genet.* **110**, 1549–1563 (2023).
45. Pott, J. et al. Genetic Association Study of Eight Steroid Hormones and Implications for Sexual Dimorphism of Coronary Artery Disease. *J. Clin. Endocrinol. Metab.* **104**, 5008–5023 (2019).
46. Neale, B. <http://www.nealelab.is/uk-biobank> (2018).
47. Wang, H. et al. Similar genetic architecture of Alzheimer's disease and differential APOE effect between sexes. *Front. Aging Neurosci.* **13**, 674318 (2021).
48. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
49. Day, F. R. et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
50. Perry, J. R. B. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
51. Oppenheimer, H. et al. Investigating Shared cardiovascular factors and genetic overlap of pregnancy-related disorders, major depressive disorder, and Alzheimer's disease. *Biol. Psychiatry* (2025).
52. Elsworth, B. MRC-IEU Consortium. <https://gwas.mrcieu.ac.uk> (2025).
53. Blokland, G. A. M. et al. Sex-dependent shared and non-shared genetic architecture across mood and psychotic disorders. *Biol. Psychiatry* **91**, 102–117 (2022).
54. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
55. Alfaro-Almagro, F. et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
56. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).
57. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).
58. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
59. Rosen, A. F. G. et al. Quantitative assessment of structural image quality. *NeuroImage* **169**, 407–418 (2018).
60. Korbmacher, M. et al. Brain asymmetries from mid- to late life and hemispheric brain age. *Nat. Commun.* **15**, 956 (2024).
61. Maximov, I. I., Alnæs, D. & Westlye, L. T. Towards an optimised processing pipeline for diffusion magnetic resonance imaging data: Effects of artefact corrections on diffusion metrics and their age associations in UK Biobank. *Hum. Brain Mapp.* **40**, 4146–4162 (2019).
62. Maximov, I. I. et al. Fast quality control method for derived diffusion Metrics (YTRIUM) in big data analysis: U.K. Biobank 18,608 example. *Hum. Brain Mapp.* **42**, 3141–3155 (2021).
63. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, San Francisco California USA, 2016).
64. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
65. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015-0047–8 (2015).
66. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
67. The GTEx Consortium et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
68. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
69. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *ELife* **7**, e34408 (2018).
70. Yavorska, O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
71. Mounier, N. & Kutalik, Z. Bias correction for inverse variance weighting Mendelian randomization. *Genet. Epidemiol.* **47**, 314–331 (2023).
72. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat. Med.* **40**, 5434–5452 (2021).
73. Oppenheimer, H. et al. No causal links between estradiol and female's brain and mental health using Mendelian randomization

(this paper). *GitHub* <https://doi.org/10.5281/zenodo.17099458> (2025).

Acknowledgements

This research has been conducted using the UKB under Application 27412. UKB has received ethics approval from the National Health Service National Research Ethics Service (ref 11/NW/O382). The study dataset complies with the Helsinki Declaration, with informed consent obtained from all participants. 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). We are grateful to all the participating families in Norway who take part in this ongoing cohort study. For generating high-quality genomic data, we thank the Norwegian Institute of Public Health (NIPH), the HARVEST collaboration, the NORMENT Centre at the University of Oslo, the Centre for Diabetes Research at the University of Bergen, deCODE Genetics, the Research Council of Norway, the South-Eastern and Western Norway Regional Health Authorities, the ERC AdG, Stiftelsen KG Jebsen, the Tron Mohn Foundation, and the Novo Nordisk Foundation. The work was performed on the Service for Sensitive Data (TSD) platform, owned by the University of Oslo, operated, and developed by the TSD service group at the University of Oslo IT-Department (USIT). Computations were also performed using resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway. While working on this study, the authors received funding from the Research Council of Norway: 326813 (A.S.), 223273 (L.T.W.), 249795 (L.T.W.), 273345 (L.T.W.), 298646 (L.T.W.), 300768 (L.T.W.); the South-Eastern Norway Regional Health Authority: 2023037 (C.B.), 2022103 (C.B.), 2018076 (L.T.W.), 2019101 (L.T.W.); the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme: 802998 (L.T.W.); the European Research Council under the European Union's Horizon Europe Research and Innovation Programme: 101163765 (C.B.); and the Swiss National Science Foundation: PZ00P3_193658 (A.-M.G.d.L.). In addition, D.v.d.M. was funded by the Research Council of Norway via a grant obtained by Prof. Ole A. Andreassen: 324252 (O.A.A.). This work was initiated while L.S.S. and A.-M.G.d.L. were affiliated with the Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne, Switzerland. We thank Prof. Ivan I. Maximov and Dr. Max Korbmaier for establishing the advanced diffusion-weighted imaging preprocessing infrastructure used for brain age prediction. Furthermore, we would like to thank Dr. Gabriëlla Blokland, Dr. Chi-Hua Chen, Dr. Hao Wang, and Dr. Cameron B. Haas for sharing their sex-specific summary statistics with us.

Author contributions

H.O.: Visualisation, Methodology, Formal analysis, Conceptualisation, Writing – review & editing, Writing – original draft. **D.v.d.M.:** Methodology, Conceptualisation, Formal analysis, Writing – review & editing. **L.S.S.:** Methodology, Conceptualisation, Writing – review & editing.

A.C.: Methodology, Writing – review & editing. **A.S.:** Formal analysis, Writing – review & editing. **O.A.A.:** Funding acquisition, Writing – review & editing. **L.T.W.:** Supervision, Conceptualisation, Writing – review & editing. **A.-M.G.d.L.:** Supervision, Methodology, Formal analysis, Conceptualisation, Writing – review & editing. **C.B.:** Supervision, Funding acquisition, Methodology, Conceptualization, Writing – review & editing, Writing – original draft.

Competing interests

O.A.A. has received speaker fees from Lundbeck, Janssen, Otsuka, Lilly, and Sunovion and is a consultant to Cortechs.ai. and Precision Health. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65878-7>.

Correspondence and requests for materials should be addressed to Hannah Oppenheimer or Claudia Barth.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Hannah Oppenheimer^{1,2} ✉, **Dennis van der Meer**³, **Louise S. Schindler**^{4,5}, **Arielle Crestol**^{1,3}, **Alexey Shadrin**³, **Ole A. Andreassen**^{3,6}, **Lars T. Westlye**^{2,3,6}, **Ann-Marie G. de Lange**^{2,4,7} & **Claudia Barth**^{1,8,9} ✉

¹Division for Mental Health and Substance Abuse, Diakonhjemmet Hospital, Oslo, Norway. ²Department of Psychology, University of Oslo, Oslo, Norway. ³Center for Precision Psychiatry, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ⁴Department of Psychiatry, University of Oxford, Oxford, UK. ⁵Social, Genetic, and Developmental Psychiatry Centre, King's College London, London, UK. ⁶KG Jebsen Center for Neurodevelopmental Disorders, University of Oslo & Oslo University Hospital, Oslo, Norway. ⁷School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. ⁸Department of Psychiatry and Neuroscience (CCM), Charité – Universitätsmedizin Berlin, Berlin, Germany. ⁹Research Unit Gender in Medicine, Charité – Universitätsmedizin Berlin, Berlin, Germany. ✉ e-mail: hannah.oppenheimer@psykologi.uio.no; claudia.barth@charite.de