Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Within-person ovarian hormone variation and structural connectome topology: a three-visit diffusion MRI study

Zhijing Chen
Guangzhou Red Cross Hospital (Guangzhou Red Cross Hospital of Jinan University)
Yiling Qiu
Guangzhou Red Cross Hospital (Guangzhou Red Cross Hospital of Jinan University)
Guiyin Chen
The Fifth Teaching Hospital of Xinjiang Medical University
Qingwen Chen
chen_qingwen0518@163.com
Liwan Central Hospital of Guangzhou
Yan Tan
Guangzhou Red Cross Hospital (Guangzhou Red Cross Hospital of Jinan University)

Article

Keywords: menstrual cycle, estradiol, progesterone, diffusion MRI, structural connectome, longitudinal analysis

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11528524/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Ovarian hormone variation has been associated with brain tissue measurements, but whether it accompanies changes in global structural network topology remains uncertain. We analysed 22 naturally cycling participants from OpenNeuro ds005360, with serum measurements and diffusion magnetic resonance imaging during menses, estimated ovulation and the mid-luteal phase. Longitudinal anatomical processing, anatomically constrained tractography and SIFT2 weighting yielded 66 connectomes with 84 nodes. Four primary mixed models tested within-person estradiol (E2) or progesterone (P4) associations with total-weight-normalized global efficiency and modularity Q, adjusting for participant-average hormone concentration, visit order and motion. No association survived four-test false-discovery-rate correction. The largest estimate was negative for E2–efficiency (standardized $\beta = -0.132$; 95% confidence interval − 0.258 to − 0.006; $p = 0.0398$; $q = 0.159$). Its participant-cluster bootstrap interval excluded zero, but permutation $q = 0.111$ also exceeded the corrected threshold. The negative direction persisted across participant omissions and alternative graph definitions, but 9 of 22 omissions gave $p \geq 0.05$. Secondary global, nodal and connected-edge analyses yielded no corrected discovery. These findings provide no multiplicity-corrected evidence of a hormone association with the tested global topology measures. They do not establish structural invariance. The E2–efficiency direction warrants independent evaluation with larger samples and denser longitudinal measurements.

Introduction

Menstrual-cycle hormone fluctuations provide a recurring source of biological variation within individuals. Their timing and magnitude differ across people, so a menstrual-phase label is an imperfect substitute for a measured endocrine exposure. Recommendations for menstrual-cycle research therefore emphasize biologically informed scheduling, direct hormone measurement and clear reporting of reproductive characteristics [1,2]. In repeated-measures neuroimaging, an additional distinction matters: an association between individuals with different average hormone concentrations need not describe what happens when an individual's hormone concentration changes. Longitudinal analyses should identify which of these relationships they estimate.

Evidence of menstrual-cycle brain variation spans several measurement scales. A multimodal systematic review documented findings in structural and functional imaging alongside substantial methodological heterogeneity [3]. More recently, dense individual sampling has described structural trajectories across different hormonal milieus [4]. In the Hormone Health Study, Rizor and colleagues associated hormone concentrations with diffusion and morphometric measurements in 30 participants [5]. Porcu and colleagues subsequently reported phase-related quantitative-anisotropy differences using differential tractography in the publicly shared 22-participant subset [6]. These observations motivate a network-level question, but tissue-sensitive and tract-specific measurements do not directly establish a change in whole-brain structural topology. The present study reuses the public subset and therefore extends an existing cohort rather than independently replicating its earlier findings.

Functional studies provide a related, distinct rationale. Repeated fMRI measurements have associated endocrine variation with resting-state connectivity [7,8]. The densely sampled 28andMe project examined whole-brain organization, transient communities and cerebellar networks within one participant [9–11]. A group magnetoencephalography study also related hormone concentrations to functional network topology [12]. Functional connectivity measures statistical coupling among signals; diffusion-derived structural networks represent reconstructed pathways between anatomical regions. Changes in functional coupling or tissue properties can occur without a detectable shift in the relative organization of reconstructed structural connections. The modalities therefore cannot be treated as interchangeable evidence of network reorganization.

Weighted global efficiency and modularity Q summarize complementary properties of a structural graph. Efficiency describes inverse shortest-path distances, whereas modularity describes the concentration of weights within communities. Both compress many connections into a single measurement. A small or spatially heterogeneous change may leave these summaries unchanged. Conversely, distributed weight changes could shift a global measure without producing a localized regional effect. Interpretation also depends on weight scale: multiplying every edge weight by the same factor changes inverse-weight efficiency without changing the relative weight distribution. Normalizing total undirected weight to one isolates the latter aspect, at the cost of removing uniform scaling effects. Raw total weight and density can then be examined as separate outcomes.

We asked whether within-person E2 or P4 fluctuations were associated with normalized global efficiency or modularity Q across three endocrine sampling windows. The two hormones and two outcomes defined four primary tests. Person-mean centring separated within-person from between-person hormone differences, and participant random intercepts accounted for repeated observations. We retained a two-sided primary hypothesis and a single four-test correction. Resampling, participant omissions and alternative graph definitions assessed uncertainty and sensitivity; secondary global and exploratory spatial analyses remained separate. This hierarchy kept the question focused. It also prevented a nominal association, or agreement among analyses of the same cohort, from being treated as independent confirmation.

Methods

Study design and endocrine measurements

This secondary analysis used the Brain Imaging Data Structure release [13] of the Hormone Health Study, OpenNeuro ds005360 version 1.0.1 [14]. The public release includes 22 of the 30 source-study participants, each with three visits and consent for data sharing. We included all 66 sessions. Source eligibility required ages of 18–30 years, nulliparity, relatively regular cycles and no recent hormonal or implant contraception. The source criteria permitted mild-to-moderate mood or attention-deficit disorders; the sample should not be described as uniformly free of psychiatric history. Detailed recruitment and eligibility information appears in the Supplementary Methods.

Visits targeted menses, estimated ovulation and the mid-luteal phase using the source scheduling procedures. We retained the released phase labels and modelled acquisition visit order separately. Blood was collected before MRI. E2 and P4 were measured by liquid chromatography–tandem mass spectrometry; luteinizing hormone (LH) and follicle-stimulating hormone (FSH) were measured by clinical immunoassay. Released concentrations were used without recalibration. Units were pg/mL for E2, ng/mL for P4 and mIU/mL for LH/FSH. Assay details and scheduling are provided in the Supplementary Methods. The source study received University of California, Santa Barbara Human Subjects Committee approval, protocol 40-22-0249. Participants in the source study provided written informed consent [5], and the public subset includes those who consented to open data sharing [14]. This analysis involved deidentified public data and no new participant contact. The secondary use of these lawfully obtained public data followed the public-data provisions of Article 32(1) of China's Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (2023).

MRI processing and connectome reconstruction

MRI was acquired on a Siemens 3 T Prisma scanner. T1-weighted images had 0.9-mm isotropic resolution; diffusion images had nominal 2 × 2 × 4-mm voxels. Only the four linear b-tensor runs were used for fibre-direction reconstruction. These inputs comprised one b0 and 68 diffusion-weighted volumes across five shells, b = 100–2,000 s/mm². Spherical-encoding runs were excluded because of their different acquisition geometry. The release contained neither acquired reverse-phase-encoding images nor field maps (Supplementary Methods).

Diffusion processing combined MP-PCA denoising, T1-informed Synb0-DisCo/TOPUP susceptibility correction, FSL eddy with a linear field model and N4 bias correction [15–18]. Rotated gradients were retained. FreeSurfer longitudinal processing used an unbiased within-participant anatomical template [19]. A fixed Desikan84 node definition combined 68 Desikan–Killiany cortical parcels with 16 bilateral subcortical/cerebellar regions [20]. Corrected diffusion images were registered to the corresponding longitudinal anatomy. Multi-shell multi-tissue constrained spherical deconvolution, cohort-average response functions and intensity normalization preceded tractography in MRtrix3 [21,22]. Anatomically constrained iFOD2 tracking generated 10 million accepted streamlines per session [23]. SIFT2 assigned continuous streamline weights, whose sums defined undirected node-pair weights [24]. Acquisition, registration, tracking and endpoint-assignment details are provided in the Supplementary Methods.

Graph definitions and quality control

Each primary graph was an unthresholded 84 × 84 SIFT2 matrix. Matrices were required to be finite, non-negative, symmetric and zero-diagonal. Dividing by the unique upper-triangular weight sum set total undirected weight to one. Weighted global efficiency averaged inverse shortest-path distances after converting positive weights to lengths, 1/w; unreachable pairs contributed zero [25]. These graph lengths are inverse-weight costs, not physical tract lengths. Weighted modularity used Louvain optimization at

resolution γ = 1 [26]. The highest-Q solution across 100 deterministic starts supplied the selected partition. Definitions and partition-stability checks are provided in the Supplementary Methods.

Quality control covered input integrity, anatomy, diffusion motion and outliers, registration, fibre orientations, tractography and matrix structure. Disconnected primary graphs, isolated nodes, invalid metrics or an incorrect node count were hard failures. Distributional and longitudinal cautions prompted review and sensitivity checks rather than automatic exclusion. No observation was excluded because of its hormone level or association result. Fixed-density graphs retained the strongest edges at 10%, 15%, 20% or 25% density before renormalization. A separate hormone-blind length-aware consistency mask retained edges within streamline-length strata, following the rationale of spatially informed consensus methods [27,28]. Atlas and reconstruction choices were held fixed because they affect graph measurements [29,30]. The Supplementary Methods describe the quality gates and alternative graphs.

Primary longitudinal analysis

E2 and P4 were fitted in separate models after log(1 + concentration) transformation in the stated units. For each participant, the mean transformed concentration across three visits was subtracted from each observation. These within-person deviations were standardized across all 66 sessions; participant-average hormone concentrations were standardized across the 22 participant means [31]. Outcomes and relative motion were standardized across sessions. Visit order was coded 1–3 and centred at the second visit (Supplementary Methods).

Each Gaussian random-intercept model included within-person hormone, participant-average hormone, visit order and relative motion as fixed effects. Restricted maximum likelihood supplied the standardized within-person coefficient, standard error, Wald 95% confidence interval and two-sided p value. The four E2/P4-by-efficiency/modularity tests formed one Benjamini–Hochberg (BH) family, with $q < 0.05$ defining a corrected discovery [32]. The coefficient describes an outcome change in pooled outcome SD units per pooled within-person SD of transformed hormone. It does not represent an individual-specific slope or a causal effect.

Supporting analyses comprised 2,000 participant-cluster bootstraps and 5,000 within-person hormone permutations per association. Each bootstrap sampled participants with replacement and retained their three visits. Percentile intervals were unadjusted. Permutations shuffled a participant's exposure values across visits while holding outcomes and other covariates fixed; empirical two-sided p values received separate four-test BH correction. This observational exposure-shuffling procedure assumes within-participant exchangeability relative to those covariates and was treated as supporting evidence. Residual diagnostics, all 22 participant omissions per model and previously identified quality-review exclusions assessed model sensitivity (Supplementary Methods). No diagnostic-driven model replacement was introduced.

Additional analyses and reporting hierarchy

Fixed-density models formed a 16-test BH family; length-aware models formed a four-test family. Secondary families comprised LH/FSH models, six phase contrasts, eight additional global-topology models and four raw-total-weight/density controls. Exploratory analyses tested 84 nodes separately within each of six hormone–metric families. The network-based statistic (NBS) assessed connected edge sets at $|t| \geq 3.1$ [33]. Its 5,000-permutation null retained the maximum component extent across both hormones and signs, supplying joint family-wise correction. The Supplementary Methods specify these procedures. For display only, node centroids from each session-specific parcellation were mapped to MNI305 space with that session's FreeSurfer talairach transform and averaged across the 66 sessions; no analysis used these coordinates.

The primary analysis plan was specified before the corresponding hormone inference but was not publicly preregistered. Fixed-density graphs were defined during blinded processing. Other extensions were specified after primary results but before inspection of their own associations. Random-slope and participant-fixed-effect estimators were post-primary checks. The Supplementary Methods report software versions, those checks and sensitivity analyses that could not be performed.

Use of generative artificial intelligence

OpenAI Codex assisted with manuscript restructuring and language revision, preparation of supplementary numerical tables from existing analysis results, reference formatting, submission-document preparation, and scripts used to revise figure annotations. The numerical table entries were checked against the analysis results, and the revised figure annotations were checked against the original vector geometry and scientific labels.

Results

Complete retention across the three sampling windows

All 22 participants contributed three sessions, with complete E2, P4, LH and FSH measurements (Fig. 1; Table 1). Mean age was 21.5 years (SD 3.1; range 18–28). Median E2 concentrations were 25.0 pg/mL during menses, 205.0 near estimated ovulation and 130.0 during the mid-luteal phase. P4 medians were 0.10, 0.65 and 11.00 ng/mL, respectively. These descriptive distributions were compatible with the intended sampling windows, without identifying the exact timing of ovulation at every visit.

Table 1
Cohort and session characteristics

Characteristic	Menses	Estimated ovulation	Mid-luteal	Overall
Participants, n	22	22	22	22
Sessions, n	22	22	22	66
Estradiol, pg/mL	25.0 [18.0, 30.0]	205.0 [132.5, 267.5]	130.0 [83.2, 157.5]	92.0 [27.0, 190.0]
Progesterone, ng/mL	0.10 [0.07, 0.18]	0.65 [0.54, 1.06]	11.00 [9.45, 17.25]	0.61 [0.12, 8.40]
Luteinizing hormone, mIU/mL	5.21 [2.92, 6.64]	47.70 [29.75, 68.77]	4.04 [2.33, 8.21]	6.61 [3.45, 26.19]
Follicle-stimulating hormone, mIU/mL	5.52 [4.39, 7.51]	11.25 [7.42, 14.50]	2.70 [2.13, 3.53]	5.40 [2.87, 8.45]
Relative motion, mm	0.120 [0.092, 0.145]	0.145 [0.120, 0.320]	0.155 [0.110, 0.200]	0.130 [0.110, 0.230]
Observed density	0.585 [0.561, 0.601]	0.592 [0.552, 0.618]	0.576 [0.560, 0.604]	0.585 [0.559, 0.608]
Normalized global efficiency	0.000943 [0.000937, 0.000985]	0.000947 [0.000937, 0.000973]	0.000954 [0.000929, 0.000973]	0.000947 [0.000935, 0.000975]
Modularity Q	0.622 [0.615, 0.628]	0.622 [0.616, 0.629]	0.625 [0.617, 0.629]	0.623 [0.616, 0.629]

Values are median [interquartile range] unless stated otherwise. Mean age was 21.5 years (SD 3.1; range 18–28). Participants recur across phases; the overall participant count is 22.

Tractography and connectome reconstruction were completed for all 66 sessions. All 66 primary graphs were connected and had no isolated nodes. Median density was 0.585, normalized global efficiency 0.000947 and modularity Q 0.623. Final graph quality control classified 38 sessions as PASS and 28 as PASS WITH CAUTION; none failed or was excluded. Median relative motion was 0.130 mm. Participant trajectories and quality-control cautions are shown in Supplementary Figure S1.

No corrected primary hormone–topology association

All four primary random-intercept models converged without optimizer warnings. None met the four-test BH threshold (Fig. 2a; Table 2). E2–efficiency had the largest absolute estimate: $\beta = -0.132$ (SE 0.064; 95% CI − 0.258 to − 0.006), with nominal $p = 0.0398$ and $q = 0.159$. The corresponding raw-outcome coefficient was -4.80×10^{-6} efficiency units per pooled within-person SD of transformed E2.

Table 2
Primary within-person hormone associations

Association	β [Wald 95% CI]	Wald p	Wald q	Bootstrap 95% CI	Perm. q
E2 / efficiency	-0.132 [− 0.258 to − 0.006]	0.0398	0.159	−0.220 to − 0.016	0.111
P4 / efficiency	-0.105 [− 0.233 to 0.022]	0.106	0.212	−0.225 to 0.018	0.136
E2 / modularity Q	0.041 [− 0.063 to 0.145]	0.442	0.442	−0.062 to 0.176	0.516
P4 / modularity Q	0.063 [− 0.039 to 0.165]	0.223	0.298	−0.041 to 0.160	0.291

Coefficients are standardized within-person estimates, adjusted for participant-average transformed hormone, visit order and relative motion, with participant random intercepts (22 participants; 66 sessions). Wald and permutation q values use separate BH corrections across four tests. Bootstrap intervals are unadjusted, using 2,000 participant resamples; permutation results use 5,000 within-person shuffles. Primary-model diagnostics and intraclass correlations are reported in Supplementary Table S2.

P4–efficiency was β = −0.105 (95% CI − 0.233 to 0.022; q = 0.212). The modularity coefficients were positive: E2, β = 0.041 (95% CI − 0.063 to 0.145; q = 0.442), and P4, β = 0.063 (95% CI − 0.039 to 0.165; q = 0.298). Their intervals included zero. The session-level within-person deviations behind all four estimates are shown in Fig. 2c. Conditional model ICCs ranged from 0.749 to 0.847, describing the fitted between-participant variance share rather than test–retest reliability.

Resampling and influence qualify the E2–efficiency direction

All 8,000 primary bootstrap fits and 20,000 permutation fits were valid. The E2–efficiency bootstrap interval was − 0.220 to − 0.016; 98.65% of its estimates were negative. Its permutation p was 0.0278, giving q = 0.111 after the separate four-test correction. The other bootstrap intervals crossed zero, and their empirical q values were 0.136–0.516. Full distributions appear in Supplementary Figure S2. These checks did not produce a corrected primary discovery.

E2–efficiency remained negative after each participant omission, with β ranging from − 0.159 to − 0.101 (Fig. 2b). Omitting sub-27 produced the greatest attenuation: β = −0.101, 95% CI − 0.237 to 0.034. Across omissions, nominal p ranged from 0.0117 to 0.142, and 9 of 22 omissions gave p ≥ 0.05. Excluding only sub-06/ses-03, the pilot session red-flagged for eddy-correction motion and assigned a predeclared exclusion analysis, left the estimate unchanged to three decimals (β = −0.132; 65 sessions). By contrast, the E2–modularity coefficient changed sign under participant omission. All 88 omission fits and 24 targeted primary sensitivity fits were valid (Supplementary Table S1).

Primary design matrices showed low collinearity: condition numbers were 1.37–1.38 and maximum VIF was 1.053. Breusch–Pagan p values were 0.145–0.961. Residual-tail cautions remained: Jarque–Bera

tests flagged three models, and maximum absolute standardized residuals were 3.42–3.62 (Supplementary Figure S3; Supplementary Table S2). No residual-based observation removal was introduced.

Graph-definition sensitivity retained direction without corrected support

At 10%, 15%, 20% and 25% density, E2–efficiency coefficients ranged from − 0.131 to − 0.128. All four had q = 0.187 within the 16-test family (Fig. 3; Supplementary Table S3). Thirteen graphs were disconnected at 10% density; all were connected at the denser thresholds (Supplementary Figure S4). The length-aware mask retained 2,035 of 3,486 possible edges, and all masked session graphs were connected. Its E2–efficiency estimate was β = −0.132 (95% CI − 0.258 to − 0.006; q = 0.157). Neither alternative-graph family produced a corrected association. The negative E2–efficiency direction also persisted in the participant-omission analyses using alternative graph definitions.

Secondary and spatial findings remained unconfirmed

No LH/FSH, phase-contrast, secondary-topology or total-weight/density family yielded a corrected discovery (Table 3; Supplementary Figure S5; Supplementary Tables S4–S5). For efficiency, estimated ovulation versus menses was β = −0.255 (95% CI − 0.578 to 0.067; q = 0.363), and mid-luteal versus menses was β = −0.257 (95% CI − 0.567 to 0.053; q = 0.363). These contrasts were adjusted for visit order and motion. They are distinct from unadjusted phase medians and from continuous hormone associations.

Table 3
Declared inference families

Family	Tests / components	Correction	Corrected findings
Primary	4	BH across 4	0
Fixed density	16	BH across 16	0
Length-aware mask	4	BH across 4	0
LH / FSH	4	BH across 4	0
Phase contrasts	6	BH across 6	0
Additional global topology	8	BH across 8	0
Weight / density controls	4	BH across 4	0
Node models	6 × 84	BH within each 84-test family	0
NBS candidate components	12	Joint permutation FWER	0

BH, Benjamini–Hochberg false-discovery-rate control; FWER, family-wise error rate; NBS, network-based statistic. The discovery threshold is 0.05 under the stated correction. Families are separate, and sensitivity results are not independent replication. Post-primary estimator checks are reported in Supplementary Table S8.

All 504 node models converged, with no $q < 0.05$ within any 84-test hormone–metric family (Fig. 4a,b; Supplementary Table S6). The minimum q was 0.0776 in the E2–nodal-efficiency family. NBS identified 12 candidate components; none survived joint family-wise correction. The largest contained five E2-positive edges among six nodes, with corrected $p = 0.216$ (Fig. 4c,d; Supplementary Table S7). Neither result localized a confirmed anatomical correlate.

Post-primary participant-fixed-effect models retained a negative E2–efficiency coefficient ($\beta = -0.134$; separate-family $q = 0.112$). The random-slope estimate was $\beta = -0.100$ ($q = 0.401$), with singular covariance and a non-positive-definite Hessian warning. Neither estimator family changed the corrected conclusion. Estimates and uncertainty for all eight post-primary models are provided in Supplementary Table S8.

Discussion

This three-visit structural-connectome study found no multiplicity-corrected association between within-person E2 or P4 variation and normalized global efficiency or modularity Q. The clearest directional estimate linked higher within-person E2 to lower efficiency. Its negative sign persisted across participant omissions and alternative graph definitions, but both primary correction procedures remained non-

significant, and nominal support depended partly on participant influence. No secondary global or exploratory spatial family supplied a corrected finding. These results limit what the measurements support, but they do not show that structural networks are invariant across the menstrual cycle.

Interpreting this result requires separating tissue variation from relative network topology. Rizor and colleagues reported hormone-associated diffusion and morphometric variation in the source cohort [5]. Porcu and colleagues examined phase-associated quantitative anisotropy in the same public subset [6]. Those outcomes need not track normalized shortest paths or community structure. Tissue-sensitive changes could affect several connections without materially changing the global summaries, or their effects could differ across regions and participants. These are possible explanations, not mechanisms demonstrated here. Because the public subset was reused, agreement with either study would still be evidence from overlapping participants. The present contribution is a different measurement and longitudinal estimand applied to that cohort, not independent replication of hormone-related brain plasticity.

Dense individual sampling approaches the same question at a different scale. Heller and colleagues examined four female series spanning two typical cycles, an endometriosis cycle and an oral-contraceptive regimen [4]. Such designs can characterize temporal and individual structure that three visits cannot resolve, but they answer a different question from a population-average association in a repeatedly assessed cohort. The two designs therefore motivate testing whether a hormone-related pattern is shared across people, specific to individuals or detectable only at certain temporal and imaging scales. Our null corrected family cannot distinguish among those possibilities.

Normalization further bounds the biological interpretation. Because each graph had unit total undirected weight, efficiency indexed the relative arrangement of reconstructed weights, rather than their overall magnitude. A uniform scaling effect was removed by design. The negative E2 coefficient, if independently reproduced, would indicate lower inverse-shortest-path efficiency under this representation; it would not show axonal loss or impaired cognition [25,34]. Separate raw-total-weight and density analyses provided no corrected evidence for those outcomes either. Modularity coefficients likewise cannot be interpreted as increased biological segregation on the basis of their positive signs alone. Nor do the opposite efficiency and modularity point estimates establish an integration–segregation trade-off: no direct test of coordinated change was performed.

The E2–efficiency estimate is useful chiefly as a quantitative target for replication. Its Wald interval and unadjusted bootstrap interval excluded zero, yet its Wald $q = 0.159$ and permutation $q = 0.111$ did not meet the declared primary criterion. Sensitivity analyses showed that the sign was not specific to one graph definition, while participant omissions exposed uncertainty in nominal support. These results describe the same participants through related analyses. They cannot multiply the amount of independent evidence or override the four-test family. Publishing the complete estimates and uncertainty makes the proposed direction testable without promoting it into a confirmed association.

The spatial analyses impose a similar boundary. Node testing covered six 84-test families, and NBS controlled a joint component search across hormones and signs [33]. No corrected result emerged. Highlighting the smallest nodal q values or drawing an apparent hormone-sensitive circuit from candidate NBS components would overstate the evidence. Conversely, these results do not rule out smaller or spatially heterogeneous changes: global summaries and heavily corrected regional tests have different sensitivities. The present data did not identify a confirmed location or connected edge set that explains the directional efficiency estimate.

Functional-network findings should also be compared at their own measurement scale. Prior studies associated endocrine variation with resting-state connectivity, network communities, cerebellar organization and magnetoencephalographic topology [7–12]. The three 28andMe reports describe complementary analyses of one participant [9–11]. Such observations neither require a parallel change in tractography-derived topology nor supply a structural mechanism for it. Directly aligned endocrine, diffusion and functional measurements would be needed to test whether the modalities covary or follow different time courses. The present results do not justify inferences about symptoms, cognitive ability or clinical risk, none of which was an outcome here.

Several design and inferential limitations remain. First, the independent participant sample was 22, not 66; three observations per person provide limited information about nonlinear responses, temporal lags or individual slopes. The narrow recruitment range and consented public subset also restrict generalization. Second, the Gaussian primary models used asymptotic Wald inference in a small cohort, and some conditional residuals had heavy tails. Participant-level resampling adds information about uncertainty but cannot compensate for the small sample. The exposure-shuffling check additionally requires exchangeability across visits despite phase-linked exposures and fixed time-varying covariates. Its empirical p values are consequently supporting sensitivity results rather than an assumption-free validation. The post-primary random-slope fits had singular covariance and a non-positive-definite Hessian, showing how little information three visits provide about between-person slope variation.

Measurement choices impose further limitations. The diffusion acquisition had anisotropic voxels and no acquired reverse-phase-encoding reference or field map, requiring synthetic-reference susceptibility correction. Conventional reconstruction used the linear-encoding subset. Atlas choice, tractography, edge weighting and thresholding all influence graph measurements [29,30]. False-positive and false-negative pathways remain possible, and streamline weights cannot be read as axon counts [34,35]. Alternative-graph checks addressed only a subset of this dependence; they did not test an independent acquisition, a second atlas or a complete alternative correction pipeline. Although acquisition times were recorded in the public dataset, scan time was not included as a covariate. Potential confounding by time of day therefore remains unassessed.

Future studies should recruit independent cohorts, sample more densely over multiple cycles and define plausible effects and equivalence margins in advance. Diffusion acquisitions with isotropic voxels and reverse-phase-encoding references, measurement-reliability assessment and prespecified comparisons

of atlases and weight definitions would help separate biological from reconstruction variation. Nonlinear or lagged analyses require sufficient temporal information and should be planned before inspecting associations. Sample-size calculations should use plausible effect ranges and reliability estimates rather than treating the present nominal coefficient as the population effect [36]. Until then, these results support a limited conclusion: no corrected global topology association was detected, and the negative E2–efficiency direction remains a specific, unconfirmed hypothesis.

Conclusion

Within-person E2 and P4 fluctuations were not associated with normalized global efficiency or modularity Q at the declared four-test FDR threshold in this 22-participant cohort. A negative E2–efficiency estimate persisted across participant omissions and alternative graph definitions but remained non-confirmatory, and 9 of 22 omissions gave $p \geq 0.05$. The data neither establish structural invariance nor support anatomical or clinical claims. Independent studies with more participants, denser sampling and planned measurement comparisons are needed to evaluate that direction.

Declarations

Code availability

The processing and analysis scripts are available from the corresponding authors upon reasonable request.

Ethics approval and consent to participate

The source study was approved by the University of California, Santa Barbara Human Subjects Committee (protocol 40-22-0249), and all participants provided written informed consent [5]. The present study used deidentified data released under a CC0 licence from participants who consented to public sharing [14]. No new participants were recruited or contacted, and no new biological samples or directly identifying information were collected. The secondary use of these lawfully obtained public data followed the public-data provisions of Article 32(1) of China's Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (2023).

Funding

This study was supported by Science and Technology Program of Guangzhou (2025A03J3375).

Acknowledgements

We thank the Hormone Health Study participants and investigators for making the data openly available.

Data Availability

The deidentified source data are openly available under CC0 from OpenNeuro, accession ds005360 version 1.0.1: https://doi.org/10.18112/openneuro.ds005360.v1.0.1. Session-level observations, derived network measurements and numerical analysis results are provided in Supplementary Data 1. Additional methods, quality-control results and sensitivity analyses are presented in the Supplementary Information.

Author contributions

Zhijing Chen: conceptualization, methodology, formal analysis, visualization, and writing of the original draft. Yiling Qiu: data curation, image processing, validation, and manuscript review and editing. Guiyin Chen: validation, interpretation of results, and manuscript review and editing. Qingwen Chen: conceptualization, methodology, supervision, and manuscript review and editing. Yan Tan: conceptualization, supervision, funding acquisition, and manuscript review and editing. All authors reviewed and approved the final manuscript.

Additional Information

Competing interests

The authors declare no competing interests.

References

1. Schmalenberger, K. M. et al. How to study the menstrual cycle: Practical tools and recommendations. *Psychoneuroendocrinology***123**, 104895 (2021). https://doi.org/10.1016/j.psyneuen.2020.104895.
2. Hampson, E. A brief guide to the menstrual cycle and oral contraceptive use for researchers in behavioral endocrinology. *Horm. Behav.***119**, 104655 (2020). https://doi.org/10.1016/j.yhbeh.2019.104655.
3. Dubol, M. et al. Neuroimaging the menstrual cycle: A multimodal systematic review. *Front. Neuroendocrinol.***60**, 100878 (2021). https://doi.org/10.1016/j.yfrne.2020.100878.
4. Heller, C. et al. Hormonal milieu influences whole-brain structural dynamics across the menstrual cycle using dense sampling in multiple individuals. *Nat. Neurosci.***28**, 2588-2600 (2025). https://doi.org/10.1038/s41593-025-02066-2.

5. Rizor, E. J. et al. Menstrual cycle-driven hormone concentrations co-fluctuate with white and gray matter architecture changes across the whole brain. *Hum. Brain Mapp.***45**, e26785 (2024). https://doi.org/10.1002/hbm.26785.
6. Porcu, M. et al. White matter microstructural changes across the menstrual cycle: a differential tractography study. *Neuroradiology***67**, 2131-2141 (2025). https://doi.org/10.1007/s00234-025-03662-6.
7. Arélin, K. et al. Progesterone mediates brain functional connectivity changes during the menstrual cycle—a pilot resting state MRI study. *Front. Neurosci.***9**, 44 (2015). https://doi.org/10.3389/fnins.2015.00044.
8. Syan, S. K. et al. Influence of endogenous estradiol, progesterone, allopregnanolone, and dehydroepiandrosterone sulfate on brain resting state functional connectivity across the menstrual cycle. *Fertil. Steril.***107**, 1246-1255.e4 (2017). https://doi.org/10.1016/j.fertnstert.2017.03.021.
9. Pritschet, L. et al. Functional reorganization of brain networks across the human menstrual cycle. *NeuroImage***220**, 117091 (2020). https://doi.org/10.1016/j.neuroimage.2020.117091.
10. Mueller, J. M. et al. Dynamic community detection reveals transient reorganization of functional brain networks across a female menstrual cycle. *Netw. Neurosci.***5**, 125-144 (2021). https://doi.org/10.1162/netn_a_00169.
11. Fitzgerald, M., Pritschet, L., Santander, T., Grafton, S. T. & Jacobs, E. G. Cerebellar network organization across the human menstrual cycle. *Sci. Rep.***10**, 20732 (2020). https://doi.org/10.1038/s41598-020-77779-4.
12. Liparoti, M. et al. Functional brain network topology across the menstrual cycle is estradiol dependent and correlates with individual well-being. *J. Neurosci. Res.***99**, 2271-2286 (2021). https://doi.org/10.1002/jnr.24898.
13. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data***3**, 160044 (2016). https://doi.org/10.1038/sdata.2016.44.
14. Rizor, E. J. et al. Hormone Health Study (HHS): Brain Anatomy and Structure (OpenNeuro ds005360, version 1.0.1). *OpenNeuro* https://doi.org/10.18112/openneuro.ds005360.v1.0.1 (2024).
15. Veraart, J. et al. Denoising of diffusion MRI using random matrix theory. *NeuroImage***142**, 394-406 (2016). https://doi.org/10.1016/j.neuroimage.2016.08.016.
16. Schilling, K. G. et al. Synthesized b0 for diffusion distortion correction (Synb0-DisCo). *Magn. Reson. Imaging***64**, 62-70 (2019). https://doi.org/10.1016/j.mri.2019.05.008.
17. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage***125**, 1063-1078 (2016). https://doi.org/10.1016/j.neuroimage.2015.10.019.
18. Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging***29**, 1310-1320 (2010). https://doi.org/10.1109/TMI.2010.2046908.

19. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage***61**, 1402-1418 (2012). https://doi.org/10.1016/j.neuroimage.2012.02.084.
20. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage***31**, 968-980 (2006). https://doi.org/10.1016/j.neuroimage.2006.01.021.
21. Jeurissen, B., Tournier, J. D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage***103**, 411-426 (2014). https://doi.org/10.1016/j.neuroimage.2014.07.061.
22. Tournier, J. D. et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage***202**, 116137 (2019). https://doi.org/10.1016/j.neuroimage.2019.116137.
23. Smith, R. E., Tournier, J. D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage***62**, 1924-1938 (2012). https://doi.org/10.1016/j.neuroimage.2012.06.005.
24. Smith, R. E., Tournier, J. D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage***119**, 338-351 (2015). https://doi.org/10.1016/j.neuroimage.2015.06.092.
25. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage***52**, 1059-1069 (2010). https://doi.org/10.1016/j.neuroimage.2009.10.003.
26. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.***2008**, P10008 (2008). https://doi.org/10.1088/1742-5468/2008/10/P10008.
27. Betzel, R. F., Griffa, A., Hagmann, P. & Mišić, B. Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Netw. Neurosci.***3**, 475-496 (2019). https://doi.org/10.1162/netn_a_00075.
28. Roberts, J. A., Perry, A., Roberts, G., Mitchell, P. B. & Breakspear, M. Consistency-based thresholding of the human connectome. *NeuroImage***145**, 118-129 (2017). https://doi.org/10.1016/j.neuroimage.2016.09.053.
29. Zalesky, A. et al. Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage***50**, 970-983 (2010). https://doi.org/10.1016/j.neuroimage.2009.12.027.
30. Roine, T. et al. Reproducibility and intercorrelation of graph theoretical measures in structural brain connectivity networks. *Med. Image Anal.***52**, 56-67 (2019). https://doi.org/10.1016/j.media.2018.10.009.
31. Curran, P. J. & Bauer, D. J. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.***62**, 583-619 (2011). https://doi.org/10.1146/annurev.psych.093008.100356.

32. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B***57**, 289-300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
33. Zalesky, A., Fornito, A. & Bullmore, E. T. Network-based statistic: identifying differences in brain networks. *NeuroImage***53**, 1197-1207 (2010). https://doi.org/10.1016/j.neuroimage.2010.06.041.
34. Jones, D. K., Knösche, T. R. & Turner, R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *NeuroImage***73**, 239-254 (2013). https://doi.org/10.1016/j.neuroimage.2012.06.081.
35. Maier-Hein, K. H. et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.***8**, 1349 (2017). https://doi.org/10.1038/s41467-017-01285-x.
36. Button, K. S. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.***14**, 365-376 (2013). https://doi.org/10.1038/nrn3475.

Figures

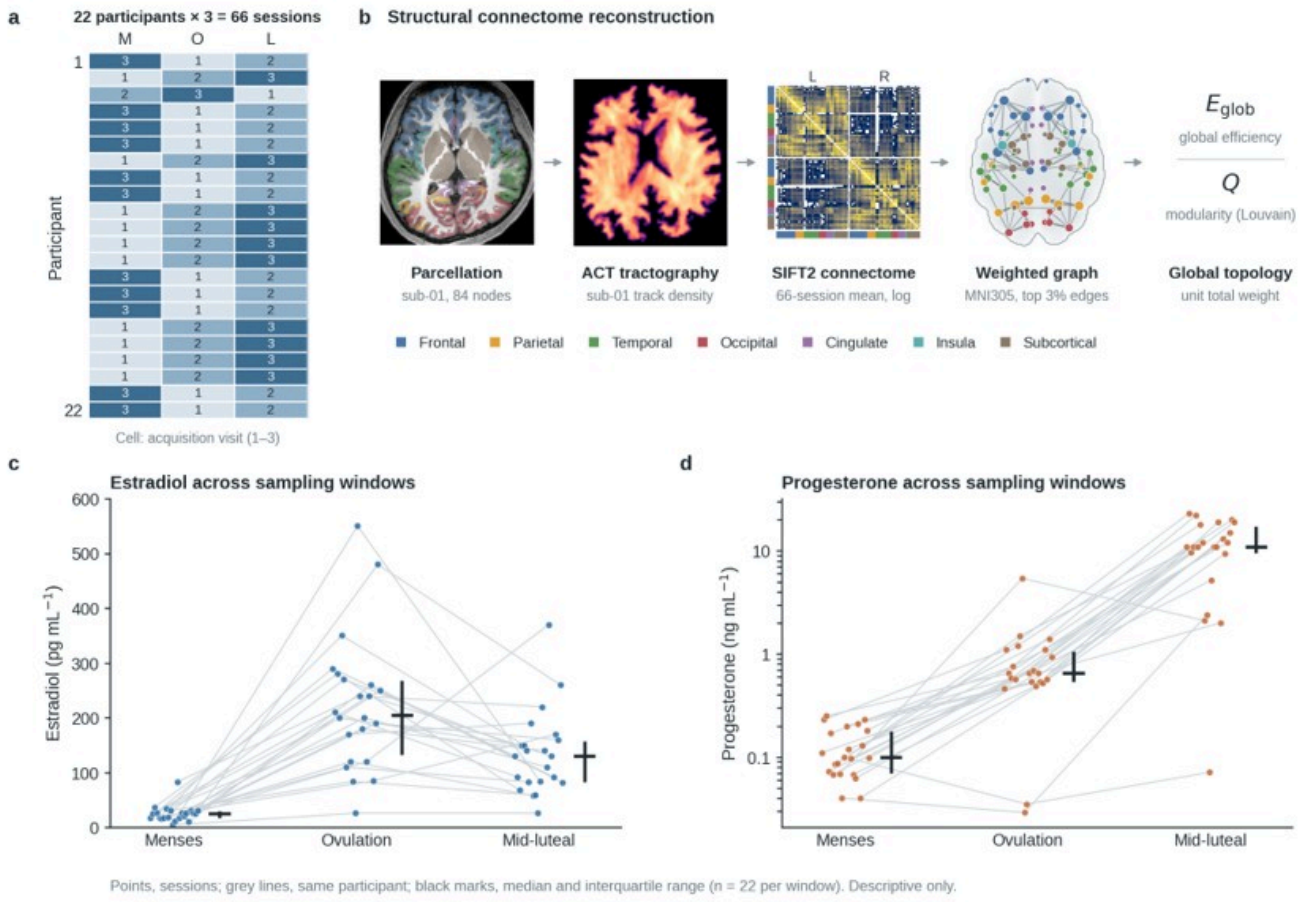


Figure 1

Study design, connectome reconstruction and observed endocrine variation. a, Design matrix for the 22 participants (rows) and three estimated endocrine windows (columns): M, menses; O, estimated

ovulation; L, mid-luteal. Each cell gives the acquisition visit (1–3) of that session, giving 66 sessions. Acquisition order varied across participants and was included separately in the models. **b**, Reconstruction and primary-outcome framework. The parcellation tile (T1-weighted image with Desikan-84 labels coloured by lobe) and the axial track-density tile are from one illustrative session (sub-01, session 01). The matrix is the mean of the 66 unit-total-weight SIFT2 connectomes on a log10 colour scale, with nodes ordered by hemisphere and lobe; blank cells are zero-weight edges. The graph shows the strongest 3% of these group-mean edges between Desikan-84 centroids in MNI305 space; node size scales with group-mean strength. Global efficiency (E_glob) and Louvain modularity Q were computed for each session on the unthresholded graph normalized to unit total undirected weight. ACT, anatomically constrained tractography; SIFT2, spherical-deconvolution informed filtering of tractograms 2. **c,d**, Released E2 (linear axis) and P4 (logarithmic axis) concentrations, with 22 session observations in each window. Lines join the same participant; small deterministic horizontal offsets separate observations. Black marks show the median (horizontal) and interquartile range (vertical). These panels are descriptive and contain no phase-comparison significance tests. E2, estradiol; P4, progesterone.

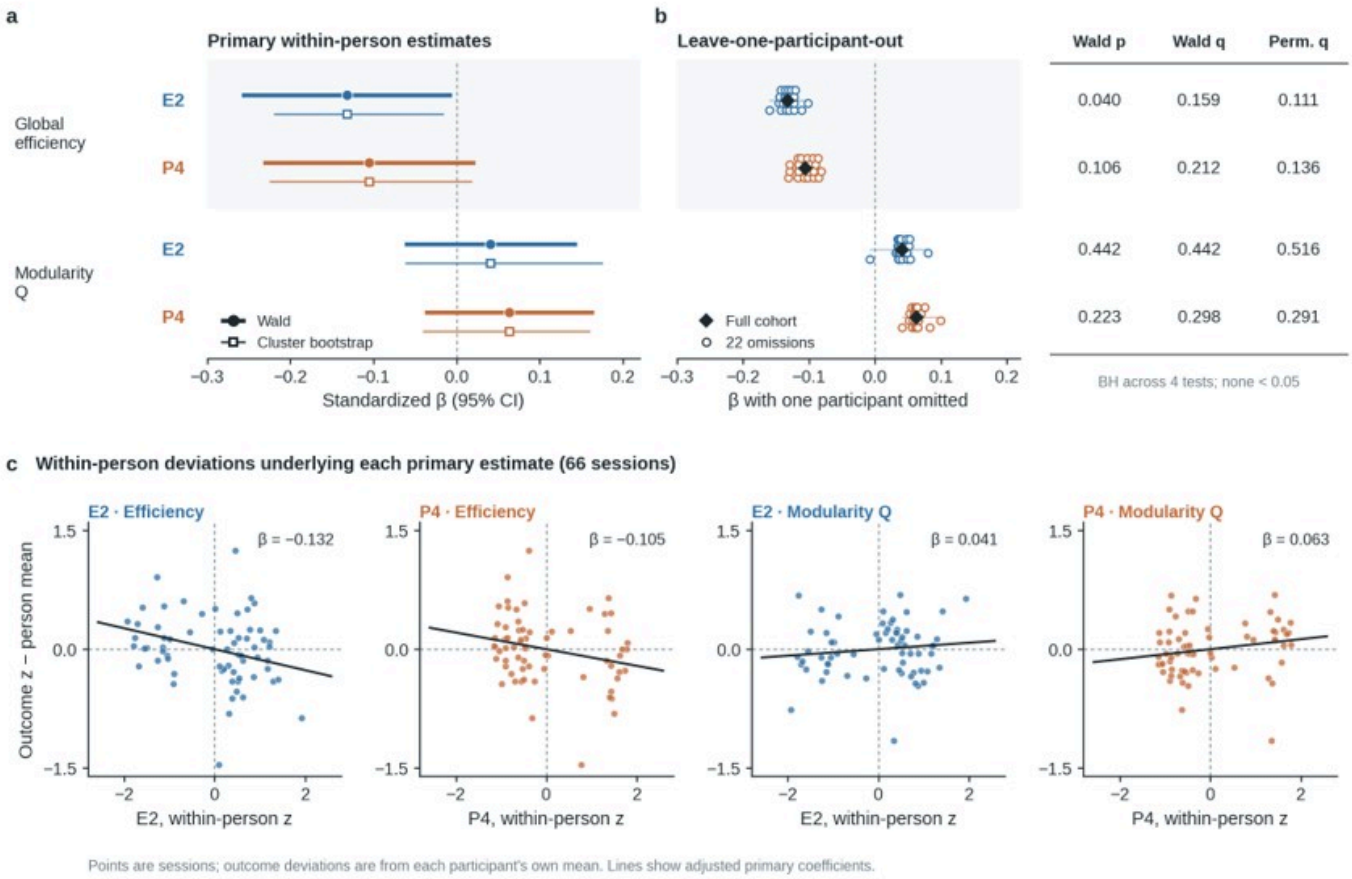


Figure 2

Primary hormone–topology estimates, participant influence and session-level data. a, Four standardized within-person coefficients from separate E2 or P4 random-intercept models for global efficiency and modularity Q. Filled circles and thicker lines show primary estimates and two-sided Wald

95% confidence intervals. Open squares and thinner lines show percentile 95% intervals from 2,000 participant-cluster bootstraps per model; squares are drawn at the primary coefficient, not the bootstrap median. Models adjust for participant-average transformed hormone, acquisition visit order and relative motion. **b**, All 22 leave-one-participant-out coefficients per association (open circles); each omission retains 21 participants and 63 sessions. Deterministic vertical offsets separate marks, the thin line spans the omission range and diamonds show the full-cohort coefficient. Panels a and b share the coefficient scale. The right-hand columns report nominal Wald p, BH-adjusted Wald q and BH-adjusted within-person permutation q from 5,000 permutations per model. Wald and permutation values were corrected separately across the same four associations; bootstrap intervals are unadjusted. **c**, Session-level data underlying each primary estimate; each panel shows all 66 sessions. The horizontal axis is the model's within-person exposure: transformed hormone minus the participant mean, divided by the pooled within-person standard deviation. The vertical axis is the standardized outcome minus the participant's own mean. Each line passes through the origin with slope equal to the covariate-adjusted primary β. The lines are visual references, not fits to the plotted points, and do not represent separate regressions. E2 is blue and P4 brown. No primary association has q < 0.05. The independent sampling units are 22 participants, each with three sessions.

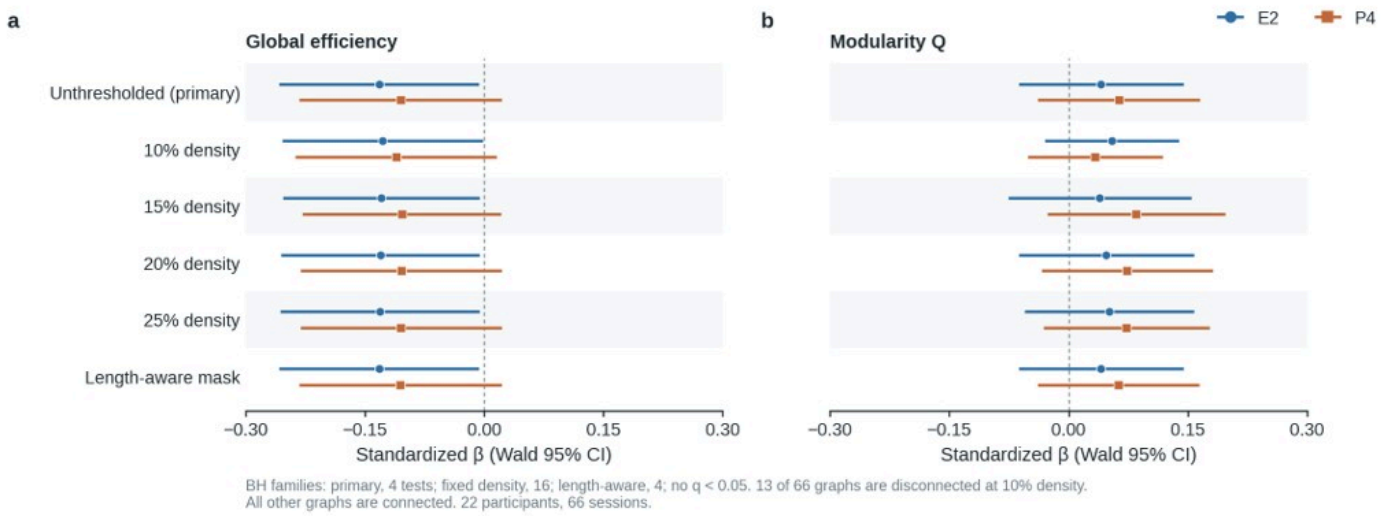
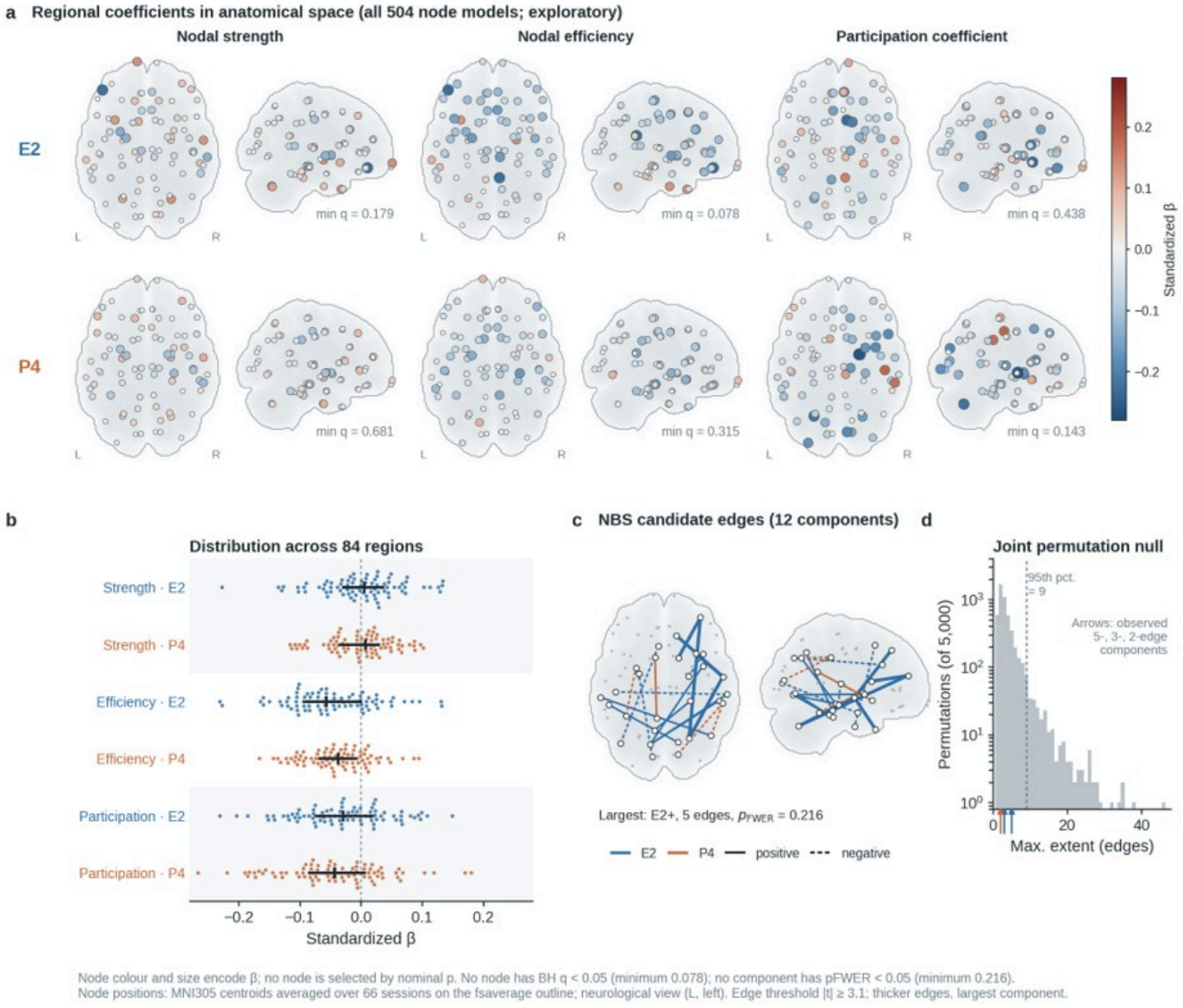


Figure 3

Sensitivity to graph definition. Standardized within-person E2 and P4 estimates for global efficiency (**a**) and modularity Q (**b**) across the primary unthresholded graph, four fixed densities and a length-aware consistency mask. Circles identify E2 and squares P4; horizontal lines are two-sided Wald 95% confidence intervals throughout. Alternating row shading is for legibility only. All models use 22 participants and 66 sessions and the primary covariate specification. Primary, fixed-density and length-aware results belong to separate BH families of 4, 16 and 4 tests, respectively; none yields q < 0.05. Thirteen session graphs are disconnected at 10% density; their unreachable node pairs contribute zero to efficiency. All denser and length-aware graphs are connected. The consistency mask retains 2,035 possible edges but does not imply that every retained edge is present in every session. Alternative

estimates are sensitivity analyses of the same cohort. Numerical estimates are provided in Supplementary Table S3.



Node colour and size encode β; no node is selected by nominal p. No node has BH q < 0.05 (minimum 0.078); no component has pFWER < 0.05 (minimum 0.216).
Node positions: MNI305 centroids averaged over 66 sessions on the fsaverage outline; neurological view (L, left). Edge threshold |t| ≥ 3.1; thicker edges, largest component.

Figure 4

Exploratory node and connected-edge results. a, All 504 standardized within-person node coefficients for three nodal metrics (columns) and two hormones (rows), shown as axial (left) and sagittal (right; anterior to the right) glass-brain projections. Node colour and area both encode the coefficient on a shared diverging scale; no node is selected or highlighted by nominal significance. Node positions are Desikan-84 centroids mapped to MNI305 with each session's FreeSurfer talairach transform and averaged over the 66 parcellations (median 2.9 mm from fsaverage atlas centroids); the outline is fsaverage. Nodes are drawn inferior to superior (axial) and lateral to medial (sagittal). Each pair shows its family's minimum BH q (84 tests). Neurological convention. **b**, The same coefficients as distributions; each point is a region, with deterministic vertical offsets. Black marks show the across-region median

and interquartile range, not confidence intervals. No node has q < 0.05 (minimum q = 0.0776). Regions are correlated outcomes, not independent biological replicates; no between-distribution test is shown. **c**, All 19 edges of the 12 candidate network-based statistic (NBS) components formed at |t| ≥ 3.1, on the same projections. Colour identifies hormone; solid lines are positive and dashed lines negative components; thicker lines mark the largest component (five E2-positive edges, six nodes). White circles are component nodes. **d**, Joint null distribution of the maximum component extent across both hormones and both signs (all 5,000 permutations; logarithmic count axis). The dashed line marks its 95th percentile (9 edges). Arrows mark the observed 5-, 3- and 2-edge components; corrected p = (1 + null maxima ≥ extent)/5,001 gives 0.216, 0.549 and 0.881, and 1 for single edges. No component has pFWER < 0.05; membership does not confer edge-wise significance. All underlying models use the same 22 participants and 66 sessions. See Supplementary Tables S6–S7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryData1.xlsx