nature portfolio

Corresponding author(s): Yubin Kim, Vivek Natarajan, Hamid Palangi, Daniel McDuff

Last updated by author(s): Jun 17, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No new participant-level wearable, laboratory or survey data were collected for the three cohort analyses in this manuscript. We analyzed existing data from the Digital Wellbeing (DWB), GLOBEM and WEAR-ME studies. The source studies used wearable devices and smartphone-based data streams, including Fitbit and smartphone logging in DWB, smartphone passive sensing and Fitbit devices in GLOBEM, and Fitbit or Google Pixel Watch devices together with the Google Health Studies application and fasting laboratory testing in WEAR-ME. The blinded expert evaluation and clinician-panel review were collected through custom web-based review interfaces that presented de-identified reports or standardized candidate cards and recorded structured ratings, free-text responses and timestamps. Source-study data-collection details are provided in the original cohort publications and summarized in the manuscript.
Data analysis	Analyses were performed using the CoDaS pipeline, which combines language-model-guided planning with deterministic Python subprocesses for data loading, exploratory data analysis, feature construction, statistical testing, machine-learning evaluation, validation checks and figure generation. CoDaS used Gemini-3.1 Pro Preview for research-intensive reasoning and code generation and Gemini-3 Flash Preview for repeated lower-latency tasks. Deterministic analysis components implemented Spearman, Pearson and Kendall correlation analyses, Benjamini–Hochberg FDR correction, permutation tests, bootstrap confidence intervals, participant-level cross-validation, Ridge/logistic/ensemble modeling, inter-rater reliability analyses and figure/table generation. Exact Python and package versions were recorded in the analysis logs and are available with the code and audit summaries on reasonable request, subject to institutional, third-party and company restrictions. Analyses were run in Python 3.11 using pandas>=2.1,<3.0, NumPy<3, SciPy>=1.11,<2, scikit-learn>=1.4,<2.

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

Data

Policy information about availability of data

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

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

GLOBEM data are available from the original GLOBEM release under the terms specified by the dataset providers. WEAR-ME data are available to approved researchers through the access process described in the original WEAR-ME study. DWB participant-level data are not publicly available and cannot be redistributed by the authors because they contain sensitive health, wearable, smartphone and survey data subject to participant-consent and institutional data-use restrictions. De-identified source data underlying the main figures and tables, the clinician-review instrument, anonymized clinician ratings and benchmark output summaries will be made available to editors and reviewers on request upon publication where permitted by consent and data-use agreements.

Research involving human participants, their data, or biological material

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

Reporting on sex and gender	Sex/gender variables were used as recorded in the source datasets and were not inferred by the authors. Because the source studies used different terminology and collection instruments, Table 1 reports female, male and other/not-reported categories as source-recorded sex/gender rather than harmonizing them into a single biological or identity construct. Subgroup consistency checks used the available source-recorded categories and should not be interpreted as evidence of sex-specific or gender-specific biological effects.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were taken from the source datasets when available and were used for cohort description rather than as biological proxies. DWB and GLOBEM race/ethnicity categories are reported as recorded in the source data. Hispanic/Latino ethnicity was collected separately in DWB and GLOBEM, so percentages may sum to more than 100%. WEAR-ME participant-level race/ethnicity variables were not available in the analytic dataset; source-study distributions are reported in the original WEAR-ME publication and discussed as a limitation. No race- or ethnicity-stratified biomarker claims are made.
Population characteristics	The study analyzed three human wearable-health cohorts and two reviewer panels. DWB included 7,497 US adults with approximately four weeks of wearable and smartphone data. GLOBEM included 704 participant-wave observations from 497 undergraduate participants across four annual cohorts. WEAR-ME included 1,078 analytic participants with wearable data linked to fasting laboratory measurements. Table 1 reports age, sex/gender, race/ethnicity where available, body-mass index where available, device type, modalities, missingness and endpoint definitions. The blinded expert evaluation included 15 domain experts in medicine, biomedical data science, machine learning, bioinformatics and digital health. The clinician-panel study included 12 physicians and physicians-in-training across psychiatry, internal medicine, family medicine, endocrinology, geriatrics, pain medicine, radiology and general practice.
Recruitment	Cohort participants were recruited by the original source studies. DWB recruited US adults for a prospective observational digital well-being study and intentionally targeted demographic diversity across race, ethnicity, sex and age. GLOBEM recruited undergraduate participants at the University of Washington across annual cohorts from 2018 to 2021. WEAR-ME recruited participants remotely across the United States through the Google Health Studies application. Expert reviewers were recruited from academic and industrial research environments based on relevant biomedical, machine-learning, bioinformatics or digital-health expertise. Clinician reviewers were drawn from academic medical centers and clinical research institutions in the United States and the Republic of Korea. Potential self-selection and cohort-generalizability limitations are discussed in the manuscript.
Ethics oversight	The DWB study protocol and participant enrollment were reviewed and approved by the Institutional Review Board of the University of Oregon (MOD00000379), and participants provided informed consent before data collection. The GLOBEM source studies underwent Institutional Review Board review and approval, and participants signed informed-consent forms before participation. The WEAR-ME study was approved by Advarra IRB (Columbia, MD; Protocol Pro00074093), and participants provided informed consent and Quest HIPAA authorization through the Google Health Studies electronic consent flow.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No prospective sample-size calculation was performed because this was an exploratory secondary analysis of existing cohorts and reviewer-panel studies. The analytic cohort sizes were DWB N = 7,497, GLOBEM N = 704 participant-wave observations from 497 unique individuals, and WEAR-ME N = 1,078 after completeness filtering. Post-hoc power calculations are reported in the manuscript: DWB had >99% power for $\|\rho\| \geq 0.10$ and an 80% minimum detectable effect of $\|\rho\| = 0.036$; GLOBEM power was computed conservatively on N = 497 unique individuals and reached 80% power for $\|\rho\| \geq 0.13$; WEAR-ME reached 80% power for $\|\rho\| \geq 0.09$. The blinded expert evaluation included 15 domain experts, 34 evaluation sessions and 76 manuscript assessments. The clinician review included 12 physicians or physicians-in-training, 17 candidate cards per reviewer and 204 card-level assessments.
Data exclusions	DWB excluded 3 of 7,500 participants because of incomplete baseline assessments, yielding N = 7,497. WEAR-ME excluded 87 of 1,165 participants with incomplete wearable feature coverage, yielding N = 1,078. In GLOBEM, features with more than 70% missingness were removed and remaining values were median-imputed within sensing wave; the analysis used 704 participant-wave observations from 497 unique individuals, with PHQ-4 coverage of 65.0%. These exclusions and preprocessing choices were based on data completeness and pipeline design. The analyses were exploratory and were not preregistered.
Replication	No candidate biomarker was externally replicated in an independent cohort with an aligned endpoint instrument. Internal reproducibility was assessed through a structured validation battery including held-out participant-level confirmation, permutation testing, bootstrap stability, leave-one-out influence, subgroup consistency, method triangulation, construct-validity and construct-independence gates, CI consistency and discriminative-signal checks. DWB and GLOBEM provided only hypothesis-generating construct-level convergence for circadian-instability-related depression candidates, because the cohorts, instruments and feature operationalizations differed. One GLOBEM candidate was explicitly flagged as unstable after held-out sign reversal.
Randomization	Participants were not allocated to experimental groups; the cohort analyses were observational. Randomization was used only in analytical or review procedures: participant-level cross-validation and held-out splits were used for model evaluation, the blinded expert evaluation randomized the mapping between system identity and model labels independently for each session, and the clinician-panel review randomized candidate-card order for each reviewer.
Blinding	The analytical CoDaS pipeline was not blinded because the system supports optional expert feedback; human feedback during discovery was restricted to mechanistic interpretation and high-level analytical guidance. The blinded expert evaluation was blinded to system identity, randomized model-label mapping, study hypotheses and other reviewers' assessments. Clinician reviewers evaluated standardized candidate cards and were not given CoDaS confidence-tier labels; task order was randomized for each reviewer.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	This study was not a clinical trial and did not assign participants to an intervention. No clinical-trial registration was required. The manuscript reports observational analyses of existing wearable, survey and laboratory datasets, together with blinded expert and clinician-panel review studies.
Study protocol	No interventional clinical-trial protocol was used for this manuscript. Source cohort protocols and data-collection procedures are described in the original DWB, GLOBEM and WEAR-ME publications and summarized in the manuscript. DWB was approved by the University of Oregon IRB, GLOBEM by the University of Washington IRB, and WEAR-ME by Advarra IRB.
Data collection	DWB data were collected from US adults in a prospective observational digital-wellbeing study using wearable and smartphone data over approximately four weeks. GLOBEM data were collected from University of Washington undergraduate cohorts from 2018 to

2021 using smartphone passive sensing and Fitbit devices. WEAR-ME data were collected remotely across the United States through the Google Health Studies application, Fitbit or Google Pixel Watch devices, and fasting laboratory testing at Quest Diagnostics centers. Expert and clinician-review data were collected asynchronously through web-based review interfaces.

Outcomes

The primary cohort endpoints were PHQ-8 total score for DWB, PHQ-4 for GLOBEM and HOMA-IR as a continuous insulin-resistance endpoint for WEAR-ME, with HbA1c-based diabetes status also reported in the cohort table. Expert-evaluation outcomes included seven Likert-scale manuscript-quality dimensions, reviewer recommendation, effort-preservation score, forced ranking, safety flags and free-text comments. Clinician-panel outcomes included seven five-point translational-relevance ratings, literature-support categories and written justifications.

Plants

Seed stocks

No seed stocks, plant specimens or other plant materials were used in this study.

Novel plant genotypes

No novel plant genotypes were generated, used or analyzed in this study.

Authentication

No plant materials, seed stocks or plant genotypes were used; therefore, no plant-authentication procedures were performed.