nature portfolio

Corresponding author(s): Sarah E. Berry

Last updated by author(s): Jan 12, 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	Data from questionnaires, clinical visits and laboratory data was entered using comma delimited files, Excel spreadsheets and Microsoft access (Microsoft Office 365 2019). CGM and accelerometer data was imported from text files into the analysis pipeline.
Data analysis	All statistical analyses were performed using R (version 3.4.2, R Core Team).

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

Data

Policy information about availability of data

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

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

The study data can be released to bona fide researchers who submit a research proposal to data.papers@joinzoe.com. All proposals will be reviewed by a sub-panel of the ZOE Scientific Advisory Board within four working weeks. To protect participant privacy, individual participant clinical data are not publicly available and

cannot be deposited in public repositories. Proposals, researchers or institutions requesting data will be approved if they meet the standard criteria related to ethics, privacy and data protection regulations. Approved researchers are required to enter into a data-sharing agreement with ZOE.

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	In the ZOE PREDICT studies, sex was considered in the study design, with sex data being self-reported by participants. The source data can be disaggregated by sex, allowing for a detailed analysis of sex-specific outcomes. The overall number of participants in each cohort, with the percentage of females, is as follows: PREDICT 1 (n=616, 75% female), PREDICT 2 (n=661, 73% female), and PREDICT 3 (n=2357, 88% female). Sex differences analyses have been conducted for interstitial glucose time in range (TIR) and coefficient of variation (CV). These analyses aimed to explore potential sex-based differences in glycaemic control. Findings from these analyses will be reported separately for each sex where applicable.
Reporting on race, ethnicity, or other socially relevant groupings	We used self-reported ethnicity data, collected through participant questionnaires, to characterize the cohorts. These data were primarily used to explore differences in continuous glucose monitoring (CGM) outcomes, such as interstitial glucose time in range (TIR), coefficient of variation (CV), and other glycaemic metrics.
Population characteristics	In the ZOE PREDICT studies, several covariate-relevant participant characteristics were collected to assess their influence on the outcomes, particularly regarding CGM metrics and other cardiometabolic health markers. These characteristics include: - Demographic Information: Sex, age, and ethnicity were collected through self-report questionnaires. These variables were used to explore differences in glucose regulation and metabolic responses across diverse groups. - Anthropometric Measures: Height, weight, body mass index (BMI), and waist circumference were recorded to assess body composition and how it relates to glycaemic outcomes and other metabolic markers. - Cardiometabolic Health Markers: Fasting plasma glucose, HbA1c, total cholesterol, HDL cholesterol, triglycerides (TG), and C-peptide levels were measured in clinical visits to evaluate participants' baseline metabolic health status and its potential influence on CGM metrics. - Blood Pressure: Systolic and diastolic blood pressure readings were collected, as hypertension is an important factor in cardiometabolic health and can affect glucose regulation. - Lifestyle Factors: Physical activity energy expenditure (PAEE) was assessed via activity trackers, while sleep duration and efficiency were also collected to understand their impact on glucose metabolism. - Dietary Intake: Dietary data were captured through food frequency questionnaires (FFQ) to assess habitual intake, as well as free-living logged diet data to track food intake and nutrient levels during the study period. -Other Health Conditions: Medical history, including the presence of type 2 diabetes, smoking status, and use of medications (such as blood-pressure-lowering treatments), were also considered as covariates for their potential influence on metabolic outcomes.
Recruitment	For PREDICT 1, participants already enrolled in the TwinsUK cohort were recruited as part of studies which are included in the cohort's annual newsletter and also mentioned on their website: http://www.twinsuk.ac.uk/ Non-twins were recruited via independent recruitment agencies, project specific website and online advertising including the use of social media platforms and social media. For PREDICT 2, Recruitment for the study was advertized to the general adult US population as well as the Stanford Nutrition Research Studies Group and Massachusetts General Hospital NAFLD Patient Registry, using online recruitment materials (emails and social media adverts) and physician letters. Study advertising was targeted to achieve a study cohort ethnically reflective of the general US population. For PREDICT 3, recruitment is be conducted through the ZOE Commercial App. Individuals who have enrolled in the ZOE Commercial App are presented with a notification inviting them to participate in the PREDICT research cohort.
Ethics oversight	For PREDICT 1, ethical approval for the study was obtained in the United Kingdom from the Research Ethics Committee and Integrated Research Application System (IRAS 236407).For PREDICT 2 and 3, ethical approval was granted by the Advarra Institutional Review Board (PREDICT 2: Pro00033432 and PREDICT 3: Pro00044316).

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	PREDICT 1 + PREDICT Cardio: 1048 individuals were recruited for the UK cohort (allowing for 80% power to detect correlations r=0.15 with p<0.0001 to allow adjustment for 500 test), 100 for the US cohort to enable replication of the larger effects (80% for r=0.28 with p<0.05) , the final sample size excludes individuals who dropped out (didn't finish the clinic visit). PREDICT 2: 978 individuals (US based) to refine the PREDICT 1 algorithms and to assess feasibility of remote postprandial assessments. The ZOE PREDICT 3 Study builds on previously collated datasets to utilise and improve the ZOE PREDICT 1 machine learning algorithm that generates personalised dietary recommendations. At the time of writing the anticipated population size is at n = 150000 (combined UK and US cohorts). Here we took a fixed sample of 4562 participants who had completed the PREDICT 3 study.
Data exclusions	Participants were excluded from the PREDICT studies for a variety of reasons related to pre-existing health conditions, medication use, dietary restrictions, and data quality. Initially, individuals with conditions like inflammatory disease, cancer, gastrointestinal disorders, diabetes, cardiovascular events, or pregnancy were excluded. Those on certain medications (immunosuppressants, antibiotics, or diabetes medication) or following a vegan diet were also excluded. The full list of study specific exclusions are available on clinical trials. Study specific exclusions were made based on CGM data quality, including insufficient free-living days, time zone changes, and device malfunction. Finally, for the main analyses, participants with diabetes or impaired glucose status (based on FPG, OGTT, or HbA1c values) and those with invalid CGM data were excluded. Dietary analyses excluded days with standardized meals and implausible calorie intake values.
Replication	In-depth analyses were primarily focused on the PREDICT 1 cohort as it contains the most deeply phenotyped individuals across the three cohorts, with the most comprehensive logged diet data and minimal missing data. This allowed for a more thorough exploration of the relationships between CGM metrics, dietary factors, and cardiometabolic health markers. Results from PREDICT 2 & 3 are mostly presented in the supplement, and are best considered separate cohorts in which we sought to replicate findings from PREDICT 1.
Randomization	Given the single arm nature of the study, randomization was not applicable
Blinding	There was no control or placebo arm therefore blinding was not applicable

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ 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	NCT03479866
Study protocol	https://clinicaltrials.gov/ct2/show/record/NCT03479866
Data collection	Data was collected at St Thomas' Hospital, London, Kings College London and Massachusetts General Hospital, Boston
Outcomes	Postprandial Metabolic Responses: Variations in serum triglycerides, glucose, and insulin were observed in response to mixed-nutrient dietary challenges. CGM and Glycaemic Metrics: Time in range (TIR) and glycaemic variability were key indicators of glucose control, with TIR3.9-5.6 being particularly relevant for identifying individuals needing diabetes screening. Diet and Lifestyle Factors: Diet composition, energy intake, macronutrient balance, and physical activity influenced glycaemic metrics and cardiometabolic health. Prediabetes and Cardiovascular Risk: CGM metrics helped distinguish individuals with prediabetes, and were predictive of atherosclerotic cardiovascular disease (ASCVD) risk, insulin resistance (HOMA-IR), and liver fat probability. Dietary Habits: Meal timing, frequency of eating occasions, and macronutrient intake were associated with glucose control and

metabolic health.
Sleep and Activity: Sleep duration and efficiency, along with physical activity, influenced glycaemic control and overall metabolic health.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.