

Corresponding author(s): Peter J Meikle; Kevin Huynh

Last updated by author(s): Jan 30, 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 consisted of epidemiological (anthropometric, clinical, lifestyle and diet) and lipidomic data. This study utilized the previously collected anthropometric, health behavior and clinical data as referenced in the methods section of the manuscript. A validated Food Frequency Questionnaire (FFQ) developed by the Anti-Cancer Council of Victoria (CCV) was used to collect data on dietary intake. Lipidomic data were generated using ESI-HPLC-MS/MS (6490 triple quadrupole, QQQ) instrument (Agilent Technologies, USA).

Data analysis

Mass Hunter (B.08.00, Agilent Technologies) in Australian Diabetes, Obesity and Lifestyle Study (AusDiab) and B10.0 software for Melbourne Collaborative Cohort Study (MCCS) sub-cohort was used for HPLC-MS/MS data acquisition and processing. Data analysis was performed as described in the method section of the manuscript. R (version 4.4.1) were used to analyze the data. In particular R packages such as glmnet and caret were used for fitting ridge and LASSO models and ggplot2 for graphics. The glm package was used to fit linear regression models. Cox regression models were fitted to compute hazard ratios using age as the time scale using the survival R package (version 4.4.1).

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](#)

We used two Australia based cohorts: The Australian Diabetes, Obesity and Lifestyle Study (AusDiab) and Melbourne Collaborative Cohort Study (MCCS). The two cohorts has a wide range of demographic, lifestyle, dietary and lipidomic data. With regard to restrictions on data availability, the participant consent obtained as part of the recruitment process for the AusDiab, it is not possible to make raw and individual level data publicly available. Individual-level data will be made available for analyses that do not conflict with ongoing studies upon reasonable written request to the study lead Professor Jonathan Shaw and the AusDiab Study Committee (Email: Jonathan.Shaw@baker.edu.au). Individual-level data for the MCCS can be made available for ethically approved projects, by application to PEDIGREE@cancervic.org.au). Responses will be provided within 2 months. The complete summary statistics for the AusDiab and the MCCS are provided in the manuscript and Supplementary files.

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

We followed the sex and gender equity in research (SAGER) guideline and included the recommended information in our study. Both sexes were included in the studies. Sex information in AusDiab cohort and MCCS sub-cohort were based on self-reported questionnaire and our study involves both sexes. We only have sex information in both cohorts and the number of participants from each sex and data reported disaggregated for sex in Table 1.

Reporting on race, ethnicity, or other socially relevant groupings

While 100% of the MCCS sub-cohort participants were white, in the AusDiab 94.7% were white/European ancestry and 5.3% were Asian/other ancestry (as self-reported by the participants). We have reported this information in the manuscript (Table 1) and in the Methods section.

Population characteristics

Basic characteristics of the cohorts is described below: (1) the Australian Diabetes, Obesity and Lifestyle Study, AusDiab n= 10,339, mean (SD) age = 51.3 (14.3), mean (SD) BMI = 26.9(4.9) and 55% women. (2) Melbourne Collaborative Cohort Study (MCCS) sub-cohort, n= 2996 mean (SD) age = 56.9 (8.6), mean (SD) BMI = 26.9 (4.4) and 51.4% women. A detailed description of the population characteristics is provided including information on sex composition (Table 1 of manuscript file).

Recruitment

This study did not directly involve the recruitment of participants. Recruitment for the AusDiab is extensively described in earlier paper (Dunstan et.al, 2002). The original MCCS participants were identified from the Victorian Electoral Enrolment Register (enrollment to vote is compulsory for all adult Australian citizens) and the Melbourne metropolitan phone directory, and were invited by letter to participate.

Ethics oversight

This study did not directly involve the recruitment of participants. We used datasets from the AusDiab biobank (project grant APP1101320) approved by the Alfred Human Research Ethics Committee, Melbourne, Australia (project approval number, 41/18) and the MCCS study was approved by the Alfred Human Research Ethics Committee (approval number, 23/013). Informed consent was obtained from all participants, and studies were conducted in accordance with the ethical principles of the Declaration of Helsinki. No participant compensation was provided.

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

For the AusDiab, sample size was selected based on precision of estimates to identify a national diabetes prevalence of 7.0% (an estimation based on results of previous surveys, and the expectation that the diabetes rate had increased over time). Accounting for the clustering of the survey design, a sample size of 11,247 (~1500 per state) was predicted to provide 95% confidence intervals of 6.2–7.8, around a diabetes estimate of 7.0%. The details on sampling and sample size had been described in earlier paper (Dunstan et.al, 2002). For lipidomics a total of 10,339 participants were included after excluding some individuals (on the basis of criteria listed below). This level of precision was regarded as acceptable, and the sample size was considered highly powered. In the MCCS sub-cohort a total of 2,996 participants who had biochemical tests including the fasting plasma were include in the lipidomic analysis.

Data exclusions

We utilized all baseline fasting plasma samples from the AusDiab cohort (n = 10,339) (Table 1) after excluding samples from pregnant women (n =21), those with missing data (n= 277), technical reasons (n=19) or whose fasting plasma samples were unavailable (n=591). No data was

excluded for the MCCS sub-cohort (lipidomics).

Replication	We replicated the associations observed in the AusDiab in the MCCS sub-cohort. The models for diet quality scores were generated in a 10 fold cross validation framework splitting the discovery dataset into training and testing datasets and evaluated all the models with the hold out testing set to confirm the robustness and avoid over fitting (within the AusDiab cohort) and externally validated the models in the MCCS sub-cohort.
Randomization	The participant recruitment and data collection for both the AusDiab and MCCS were conducted before the current study. It is important to note that the objectives of the original studies and the current study differ. During data analysis, appropriate adjustments were made by including covariates such as age, sex, and BMI and other potential confounders. Additional confounders were also taken into account as deemed necessary for the analysis. Samples were randomized prior lipidomic analysis.
Blinding	Since data collection and data analysis were carried out independently by different researchers, no additional blinding measures were implemented. Data analysts, including statisticians and biotech-informaticians, worked with deidentified data, ensuring that they are unaware of any specific laboratory, clinical, or lipidomics information. All statistical analyses and graphics are conducted using computer software without any manual interventions.

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	The study is a population study rather than a clinical trial.
Study protocol	AusDiab is extensively described in earlier paper (Dunstan et.al, 2002) reference 103 in the manuscript. The Melbourne Collaborative Cohort Study (MCCS), also known as Health 2020, was planned in the late 1980s and established in the early 1990s as an omnibus cohort to investigate prospectively the roles of diet and lifestyle in causing cancer and other non-communicable diseases. MCCS protocol is detailed in the publication by Milne RL et. al, reference 105.
Data collection	Data collection for the AusDiab is extensively described in earlier paper (Dunstan et.al, 2002). The original MCCS participants were identified from the Victorian Electoral Enrollment Register (enrollment to vote is compulsory for all adult Australian citizens) and the Melbourne metropolitan phone directory, and were invited by letter to participate. Detail on data collection for MCCS can be found in Milne RL et. al, reference 105.
Outcomes	The primary outcomes in the AusDiab were diabetes and for MCCS was cancers.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A