nature portfolio

Corresponding author(s): Mingyang Song

Last updated by author(s): Dec 8, 2025

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 software was used for data collection.
Data analysis	The statistical analyses were conducted in SAS (version 9.4), and the main figures were generated in RStudio (version 2021.09.0). The computer code used to generate results in this study is publicly available on GitHub (https://github.com/gigifang925/PAPattern_ChronicDisease/tree/main) and archived on Zenodo (DOI: 10.5281/zenodo.17857981).

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 data supporting the findings from this study is available within the manuscript and its supplementary information. The raw data and processed data of the Health Professionals Follow-up Study and Nurses' Health Studies are available under restricted access for participant confidentiality and privacy. The procedures to

access the cohort data is described at https://hsph.harvard.edu/research/health-professionals/ (email: hpfs@hsph.harvard.edu) and https://nurseshealthstudy.org/researchers (email: nhsaccess@channing.harvard.edu). Applications to access the cohort data will be reviewed by our External Collaborators Committee for scientific aims, evaluation of the fit of the data for the proposed methodology, and verification that the proposed use meets the guidelines of the Ethics and Governance Framework and the consent that was provided by the participants. Responses are typically provided within approximately two weeks. Source data are provided with this paper.

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	Both self-reported sexes (males and females) were included in the study. The study population in our analyses comprised 45 426 males from the HPFS, 72 942 females from the NHS, and 113 120 females from the NHS II. The sex, number and age of participants in each cohort are described in Methods – Study population section and in Table 1. We conducted subgroup analysis by sex and the findings apply to both males and females. The effect estimates (95% confidence intervals) in males and females separately are reported in the Results section and presented in Supplementary Table 7.
Reporting on race, ethnicity, or other socially relevant groupings	We drew data from three large prospective cohorts in the US: the Health Professional Follow-up Study (HPFS) began in 1986 and enrolled 51 529 male health professionals aged 40–75 years from all 50 states; the Nurses' Health Study (NHS) began in 1976 and included 121 700 female registered nurses aged 30–55 years from 11 states; the NHS II began in 1989 and enrolled 116 429 younger female registered nurses aged 25–42 years. Self-reported race was assessed once with a questionnaire. Over 90% of participants are white (Table 1).
Population characteristics	45 426 males from the HPFS, 72 942 females from the NHS, and 113 120 females from the NHS II aged 25-75 years and free of major chronic diseases with detailed physical activity assessments at baseline were included in our analyses. Table 1 shows the age-standardized characteristics of the study population.
Recruitment	The Health Professional Follow-up Study (HPFS) began in 1986 and enrolled 51 529 male health professionals aged 40–75 years from all 50 states; the Nurses' Health Study (NHS) began in 1976 and included 121 700 female registered nurses aged 30–55 years from 11 states; the NHS II began in 1989 and enrolled 116 429 younger female registered nurses aged 25–42 years. Follow-up questionnaires were sent to participants in these cohorts to collect and update their lifestyle and medical history every two years. Health professionals who volunteer to participate in these cohorts typically have higher health literacy and greater access to care. As a result, self-selection into the cohorts and the restriction of participants to health professionals in the US may limit generalizability to more demographically and socioeconomically diverse populations. Loss to follow-up could also introduce selection bias; however, the follow-up rate of approximately 90% across all three cohorts substantially reduces potential for such bias.
Ethics oversight	The Health Professionals Follow-up Study, Nurses' Health Study, and Nurses' Health Study II were approved by the Institutional Review Board at the Brigham and Women's Hospital, the Harvard T.H. Chan School of Public Health (IRB protocol number: 1999P011114; 1999P003389; 10162). The written informed consent form was waived for the questionnaire components of these cohorts under IRB 45 CFR 46.116(f) and 46.117(c). Each questionnaire was accompanied by a cover letter stating that "your answers will be used for medical statistical purposes only." Completion and return of the questionnaires were considered implied consent. Written informed consent was obtained for medical record acquisition. The study was performed in accordance with the Declaration of Helsinki.

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	We drew data from three preexisting large cohorts. The study population in our analyses comprised 231 488 health professionals in the US. We documented 72 648 cases of major chronic diseases, including 23 034 T2Ds, 17 415 major CVDs, and 41 525 cancers, allowing us to detect the major findings of the present study with a statistical power of > 90% (post hoc analysis).
Data exclusions	Participants were excluded at baseline if they had reported a history of cancer, cardiovascular diseases or diabetes, or had missing data on age or physical activity. After exclusions, 45 426 males from HPFS, 72 942 females from NHS, and 113 120 females from NHS II aged 25-75 years were included (Supplementary Fig. 1)
Replication	This study is an observational study and we drew data from three preexisting large cohorts of health professionals in the US. Data and code availability statement has been provided. The validity and reproducibility of questionnaires and outcome ascertainment have been reported previously and described in the Methods.

Randomization	Not applicable. This study is an observational study.
Blinding	Not applicable because this is an observational study without any interventions. Study investigators who reviewed medical records to confirm disease diagnosis were blinded to the exposure information.

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

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.*