nature portfolio

Corresponding author(s):   Robbee Wedow and Andrea Ganna

Last updated by author(s):   Apr 1, 2023

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 two primary nonresponse options "I Don't Know" and "Prefer Not to Answer" along with genetic data in the form of genotypes. Details on how these response options were extracted and analyzed are available in the manuscript and Methods. Details on how genotypes were obtained in each dataset are available in the Methods section of the manuscript.
Data analysis	Data analysis was conducted as specified in the Methods section of the manuscript. Hail Version 2.0 was used for GWAS, imputations was performed with IMPUTE2, factor analyses were performed with relevant packages in R Version 4.2.3, and polygenic scores were generated with PLINK Version 1.9 and LDpred2. All code has been updated and is available at https://github.com/gianmarcomigno/Item-nonresponse.

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 <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. 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 <u>policy</u>

> The GWAS results for our PNA and IDK phenotypes are available through the GWAS catalog accession nos. GCST90266936 for PNA and GCST90266935 for IDK.

Research involving human participants, their data, or biological material

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

Reporting on sex and gender	Sex was used as a covariate in all of the analyses presented in this study.
Reporting on race, ethnicity, or other socially relevant groupings	Genetic ancestry was determined using genetic data (using Principal Components analysis). Here we limited our sample to only individuals of European ancestry due to the statistical confounds presented by population stratification, as is standard in the literature. All GWAS and polygenic prediction exercises also controlled for genetic ancestry (the top 20 principal components of the genetic variance-covariance matrix of the genetic data for GWAS analyses and the top 10 for polygenic prediction exercises).
Population characteristics	Population characteristics for both samples are described in the "Behavioural & social sciences study design" section below.
Recruitment	The UK Biobank (UKB) is a health resource which has the purpose of improving the prevention, diagnosis, and treatment of human disease75. It consists of a prospective cohort of 502,620 men and women aged 40-69 recruited in the years 2006-2010 throughout the United Kingdom. The touchscreen questionnaire is a collection of self-reported information regarding general health, dietary habits, physical activity, psychological and cognitive states, sociodemographic factors, etc. We began with 361,194 unrelated individuals of European genetic ancestry who passed quality control measures (https://www.nealelab.is/uk-biobank/ukbround2announcement). We excluded individuals who were enrolled only in the UKB pilot study (N=335). Participants who decided to terminate the touch screen questionnaire were asked to select PNA to all subsequent questions, and they were kept in our analyses. Conversely, individuals who withdrew from the study without filling out the touchscreen survey were excluded from the analysis (N=231). As a result, a total of N=360,628 participants took part in the survey and answered every question of interest in the study; this is the final analytic sample size. Add Health originated as an in-school survey of a nationally representative sample of US adolescents enrolled in grades 7 through 12 during the 1994-1995 school year. Respondents were born between 1974 and 1983, and a subset of the original Add Health respondents has been followed up with in-home interviews, which allows researchers to assess correlates of outcomes in the transition to early adulthood. In Add Health, the mean birth year of respondents is 1979 (SD = 1.8), and the mean age at the time of assessment (Wave 4) is 29.0 years (SD = 1.8). All phenotypes included in this study come from Wave 4, the latest wave of Add Health data collection (2007-2009).
Ethics oversight	Use of the UK Biobank data was approved under application 31063. Because of the sensitive nature of this study, we also explicitly sought permission for the specific scope of this paper (i.e., "to also study response rates and response characteristics (e.g., how often a response is left unanswered) and to examine whether there are any genetic factors that correlate with these response phenotypes"), which the UK Biobank granted under the same application. Analysis of the UK Biobank data was reviewed by the Partners HealthCare IRB (Partners Human Research), which determined in expedited review that the project met the US federal criteria definition of "not human subjects research" (Protocol #2019P000883 titled "Behavioral Genetics Study of Responsiveness from UKBB Questionnaires"). Analysis of the Add Health data was reviewed by the Office of Research Subject Protection (OSRP) at the Broad Institute of MIT and Harvard, which determined that the project met US federal criteria for exemption from IRB review (Project #0001 titled "Genetic and environmental factors influencing complex social behavior").

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

Behavioural & social sciences study design

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

Study description

We performed the most in-depth nvestigation of item nonresponse behavior to date. We created factor scores for the two most common nonresponse answers, "Prefer Not to Answer" (PNA) AND "I Don't Know" (IDK) in our datasets. We conducted complementary phenotypic and genetic analyses to gain insights that would not otherwise have be obtainable by solely leveraging questionnaire-based information. All data was quantitative.

Research sample

Using the UK Biobank, we examined nonresponse behavior in 109 questionnaire items for 230,629 individuals. We performed replication exercises using polygenic scores for 3,414 individuals in the Add Health study. We chose these samples because of the large and required sample size for genetic analyses (UK Biobank) and for existence of well-phenotyped replication data (Add Health). the UK Biobank is not a nationally representative study, while the Add Health study is a US-based nationally representative study.

Sampling strategy

Analytic samples were decided by using the samples of subsamples that had the largest N for a given outcome under study. This strategy was chosen, because large samples are prioritized to have enough statistical power to isolate small genetic associations. This is common practice in the field.

Data collection

The UK Biobank (UKB) is a health resource which has the purpose of improving the prevention, diagnosis, and treatment of human disease75. It consists of a prospective cohort of 502,620 men and women aged 40-69 recruited in the years 2006-2010 throughout the United Kingdom. The touchscreen questionnaire is a collection of self-reported information regarding general health, dietary habits, physical activity, psychological and cognitive states, sociodemographic factors, etc. We began with 361,194 unrelated individuals of European genetic ancestry who passed quality control measures (https://www.nealelab.is/uk-biobank/ ukbround2announcement). We excluded individuals who were enrolled only in the UKB pilot study (N=335). Participants who decided to terminate the touch screen questionnaire were asked to select PNA to all subsequent questions, and they were kept in our analyses. Conversely, individuals who withdrew from the study without filling out the touchscreen survey were excluded from the analysis (N=231). As a result, a total of N=360,628 participants took part in the survey and answered every question of interest in the study; this is the final analytic sample size.

Add Health originated as an in-school survey of a nationally representative sample of US adolescents enrolled in grades 7 through 12 during the 1994-1995 school year. Respondents were born between 1974 and 1983, and a subset of the original Add Health respondents has been followed up with in-home interviews, which allows researchers to assess correlates of outcomes in the transition to early adulthood. In Add Health, the mean birth year of respondents is 1979 (SD = 1.8), and the mean age at the time of assessment (Wave 4) is 29.0 years (SD = 1.8). All phenotypes included in this study come from Wave 4, the latest wave of Add Health data collection (2007-2009).

Timing

The UK Biobank is a prospective cohort of 502,620 men and women aged 40-69 recruited in the years 2006-2010 throughout the United Kingdom.

Add Health originated as an in-school survey of a nationally representative sample of US adolescents enrolled in grades 7 through 12 during the 1994-1995 school year. Respondents were born between 1974 and 1983, and a subset of the original Add Health respondents has been followed up with in-home interviews, which allows researchers to assess correlates of outcomes in the transition to early adulthood.

Data exclusions

We use only individuals of European ancestry due to the statistical confounds presented by population stratification, as is standard in the literature.

Non-participation

Participants were able to select nonresponse options through the questionnaires in the UK Biobank and Add Health. We used these nonresponse options as the primary analysis variables in our study.

Randomization

Randomization was not a relevant component in this study, as there were no experimental conditions being tested and participants were divided into whether they answered questionnaires with nonresponse options or not.

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