

Searching for the causal effects of BMI in over 300 000 individuals, using Mendelian randomization

Louise A C Millard ^{1,2 *}, Neil M Davies ¹, Kate Tilling ¹, Tom R Gaunt ¹, and George Davey Smith ¹

¹ MRC Integrative Epidemiology Unit (IEU), Bristol Medical School (Population Health Sciences), University of Bristol, Bristol, UK

² Intelligent Systems Laboratory, Department of Computer Science, University of Bristol, Bristol, UK

* Email: louise.millard@bristol.ac.uk

ABSTRACT

Mendelian randomization (MR) has been used to estimate the causal effect of body mass index (BMI) on particular traits thought to be affected by BMI. However, BMI may also be a modifiable, causal risk factor for outcomes where there is no prior reason to suggest that a causal effect exists. We perform a MR phenome-wide association study (MR-pheWAS) to search for the causal effects of BMI in UK Biobank (n=334 968), using the PHESANT open-source phenome scan tool. Of the 20 461 tests performed, our MR-pheWAS identified 519 associations below a stringent P value threshold corresponding to a 5% estimated false discovery rate, including many previously identified causal effects. We also identified several novel effects, including protective effects of higher BMI on a set of psychosocial traits, identified initially in our preliminary MR-pheWAS and replicated in an independent subset of UK Biobank. Such associations need replicating in an independent sample.

INTRODUCTION

Body mass index (BMI), taken to be a general indicator of adiposity, has been associated with many traits and diseases¹⁻³. These observational associations may be due to a causal effect of adiposity (as reflected in BMI), a causal effect of the phenotype on BMI, or confounding. Mendelian randomization (MR) estimates the causal effect of an exposure on an outcome using genetic variants as an instrumental variable for the exposure^{4,5}. To date, MR has been used to assess whether BMI causally affects a vast array of phenotypes⁶⁻²¹. These results suggest that higher BMI leads to an earlier age at menarche⁹, a reduction in physical activity⁶, an increased risk of coronary heart disease²², cancer^{10,18} and asthma²¹, a lower risk of Parkinson's disease¹⁷, and changes in metabolite concentrations⁷.

Hypothesis-free searching is an established method to identify novel associations, such as genetic variants associated with a particular phenotype in genome-wide association studies (GWAS)²³. In contrast to hypothesis-driven analyses, a hypothesis-free analysis can identify novel associations where there is no prior expectation that an association might exist, and should help to avoid publication bias as all results are published, not just the most “statistically significant”. Phenome scans are a class of hypothesis-free scan that test the association of a variable of interest with a potentially large array of phenotypes – the “phenome”²⁴. Phenome scans commonly seek to identify phenotypes associated with a genetic variant (phenome-wide association studies – pheWAS) or an observed phenotype (environment-wide association studies – EnWAS). A third type of phenome scan, MR-pheWAS, uses MR to search for the causal effects of a particular exposure²⁵.

To date, only a small number of phenome-wide scans of BMI have been published. We recently published a MR-PheWAS that searched for the causal effects of BMI in circa 8000 participants in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort, across 172 outcomes²⁵. This study confirmed several known associations such as with leptin levels and blood pressure, and identified potentially novel associations, such as with a self-worth score. Cronin et al. performed a pheWAS of *FTO* genetic variants within electronic health record phenotypes, in circa 25 000 participants, and identified novel associations²⁶. For instance, a genetic predisposition to a higher BMI was associated with a higher risk of fibrocystic breast disease and non-alcoholic liver disease²⁶.

UK Biobank is a prospective cohort of circa 500 000 participants²⁷, and several hypothesis-driven MR analyses of BMI have been performed in this study, to date^{20,21,28,29}. The large sample size offsets the multiple testing burden of hypothesis-free analyses, providing an opportunity to search for causal effects with MR-pheWAS. Recently, we published an open-source tool for performing phenome scans (including MR-pheWAS) in UK Biobank – the PHEnome Scan ANalysis Tool (PHESANT)³⁰. PHESANT enables comprehensive phenome scans to be performed across a large and diverse set of phenotypes – all continuous, integer and categorical fields in UK Biobank – where previously researchers would restrict their analysis to a homogeneous subset of phenotypes that could be processed in a consistent fashion³⁰.

In this study we search for the causal effects of adiposity, using the PHESANT tool in UK Biobank. We use BMI as a surrogate measure of adiposity. In our presentation of

PHESANT³⁰ we analyzed the initial non-random sample of circa 115 000 participants that was available in UK Biobank at that time. We found that participants with a genetic propensity to a higher BMI were less likely to perceive themselves as a nervous person or to call themselves tense or ‘highly strung’³⁰. In this work, we search for the causal effects of BMI in circa 330 000 participants in UK Biobank satisfying our inclusion criteria. BMI is a well-studied phenotype, hence this study serves as a model for future MR-pheWAS that may investigate phenotypes with much weaker priors regarding their causal effects.

RESULTS

We checked the strength of the association between the BMI allele score and BMI, and found a standard deviation (SD) increase in BMI allele score was associated with a 0.63 kg/m² increase in BMI (95% confidence interval (CI): 0.62, 0.65, F statistic=6052.88).

Results of MR-pheWAS analysis

The results of our MR-pheWAS include 20 461 tests ranked by P value, given in Supplementary data file 1. Figure 2 shows the number of fields reaching each stage of the PHESANT automated pipeline. A QQ plot is given in Figure 3, and the PHESANT-viz visualization can be found at [<http://www.datamining.org.uk/PHESANT/results-bmi.html>]. We identified 519 results at a false discovery rate of 5% (using a P value threshold of $0.05 \times 519 / 20461 = 1.27 \times 10^{-3}$), given in Supplementary table 4, and of these, 259 results had a P value lower than a stringent Bonferroni corrected threshold of 2.44×10^{-6} ($0.05 / 20461$).

Our MR-pheWAS identified several known effects of BMI. For example, a genetic predisposition to a higher BMI was associated with an increased risk of diabetes¹⁹ (field ID (FID)=2443), hypertension⁴⁰ (FID=41204 value I10, and FID=4079), a higher bone mineral density⁸ (FID=3148) and an earlier age of puberty in both sexes⁹ (FIDs={2714, 2375, 2385}). Figure 4 shows that, when restricting to psychosocial traits, a subset is associated more strongly than would be expected by chance including several nervousness / anxiety traits.

PHESANT sensitivity analysis

Our tests of association of assessment centre with the BMI genetic score found that, while adjusting for the genetic principal components attenuated the association, an association remained (likelihood ratio test P values of 6.49×10^{-26} , 4.44×10^{-10} and 6.15×10^{-6} when adjusting for age and sex only, and additionally the first 10 and first 40 genetic principal components, respectively). Supplementary figure 1 compares the results of our main MR-pheWAS analysis, adjusting for age, sex and the first 10 genetic principal components, with our sensitivity analysis, additionally adjusting for assessment centre and genetic batch. Large differences between these results often occurred for categorical outcomes with small numbers in a certain category (for multinomial logistic model this can cause a zero P value when the likelihood of the model increases to a limit⁴¹; see examples in Supplementary table 5).

Detailed follow-up of potentially novel results

Our hypothesis-free scan found that a genetic predisposition to a higher BMI was associated with a decreased reporting of being a nervous person (FID=1970,

$p=2.92 \times 10^{-24}$), being tense or ‘highly strung’ (FID=1990, $p=4.47 \times 10^{-12}$), suffering from ‘nerves’ (FID=2010, $p=5.47 \times 10^{-10}$), or being a worrier (FID=1980, $p=4.42 \times 10^{-7}$). Table 1 and Supplementary figures 2 and 3 present the causal effects of BMI on these nervousness traits, estimated using two-stage instrumental variable analyses. A 1 kg/m² increase in BMI caused a 6.1% [95% CI: -7.2%, -5.0%] decrease in risk of reporting as a ‘nervous’ person. Each 1 kg/m² increase in BMI was observationally associated with a 3.6% [95% CI: -3.8%, -3.4%] decrease in risk of self-reporting as a ‘nervous’ person. Adjusting for age, sex and the first 40 genetic principal components did not meaningfully affect the results (see Supplementary table 6). Furthermore, causal estimates using the discovery subsample of UK Biobank (used in our initial presentation of PHESANT³⁰) and replication subsample (comprising additional participants used in this study) were consistent for these four nervousness/anxiety phenotypes.

As a sensitivity analysis, we estimated causal effects of BMI using two-sample MR methods. Supplementary figures 4 and 5 show the MR-Egger regression and simulation extrapolation (SIMEX) plots, respectively. We found little evidence of directional pleiotropy, as all confidence intervals for intercept estimates of MR-Egger included zero. While there was evidence that the ‘NO Measurement Error’ (NOME) assumption was not fully satisfied (I_{GX}^2 statistic = 0.893), estimates using MR-Egger with and without SIMEX were highly consistent.

Supplementary figures 6 and 7 show the smoothed empirical densities of the SNP effect estimates using mode-based estimator (MBE) smoothing parameter ϕ values of 0.5, 0.75 and 1. We choose to use $\phi=0.75$ as this provided sufficient smoothing while still

allowing multiple peaks to be captured. The effect estimates across IV methods – IV probit regression, and the two-sample approaches; MR-Egger, weighted median and MBE – were broadly consistent. We did, however, see a difference between the estimates of the IV probit regression and weighted MBE for reporting ‘being a worrier’ (FID=1980), where a 1 SD higher BMI was associated with a 12.7% [95% CI: -17.1%, -8.0%] decrease in the risk of reporting ‘being a worrier’ using IV probit regression, but a 6.5% [95% CI: -5.2%, 19.7%] increase in risk of reporting ‘being a worrier’ using the weighted MBE ($\phi=0.75$). Furthermore, the causal estimates using the 95 SNP score versus a BMI-associated SNP in *FTO* (“*FTO* SNP”) were broadly consistent for the outcomes in our follow-up analysis, except for reporting ‘being a worrier’ (FID=1980).

The results of these four nervousness/anxiety traits do not provide a form of replication for each other because these traits are highly correlated (see Supplementary table 7; all accuracy between 57.59% and 83.49%) and so are more likely to agree compared to tests using independent traits or a replication on an independent dataset.

We searched for other nervousness related traits by searching for the terms: “worr”, “nerv”, “tens”, “anxi”, in our results listing, to determine if other related traits were tested and an association not identified, to assess the strength of the evidence when considering results for all similar phenotypes. This identified three self-reported phenotypes describing: 1) whether the participant has seen a psychiatrist for nerves, anxiety, tension or depression, (FID=2100; estimated odds ratio of 1.009 [95% CI: 0.994, 1.023]), 2) whether the participant has seen doctor (GP) for nerves, anxiety, tension or depression" (FID=2090; estimated odds ratio of 1.006 [95% CI: 0.994, 1.017]), and 3) the frequency of tenseness / restlessness in last 2 weeks (FID=2070; not

at all versus several days or more; estimated odds ratio of 0.984 [95% CI: 0.973, 0.996]).

DISCUSSION

In this study we used the PHESANT phenome scan tool to search for the causal effects of BMI – a MR-pheWAS analysis – in a sample of circa 330 000 participants in UK Biobank, across over twenty thousand diverse phenotypes. This systematic approach helps to avoid biases associated with hypothesis-driven analyses, where a researcher might try several tests of association until a desired result is found, and should help to avoid publication bias as all results are published, not just the most “statistically significant”⁴².

Our results include associations consistent with previous MR studies, such as adverse effects of higher BMI on risk of diabetes¹⁹ and hypertension⁴⁰, a higher bone mineral density⁸ and an earlier age of puberty in both sexes⁹. Consistent with our preliminary results³⁰, we identified an association with a set of nervousness phenotypes, where a genetic predisposition to a higher BMI was associated with a person being less likely to call themselves a nervous person, tense or highly strung, a worrier, or to report they suffer from ‘nerves’. We followed up this analysis to estimate the causal effect and found that each 1 kg/m² decrease in BMI caused a 6.1% [95% CI: 5.0%, 7.2%] increase in likelihood of self-reporting being a nervous person. The causal estimates were more extreme than the observational estimates for three of these outcomes. Furthermore,

causal estimates using the discovery and replication subsamples of UK Biobank were consistent for these four nervousness/anxiety phenotypes.

We searched our results for other nervousness traits to determine the strength of the evidence in the context of the results of related phenotypes, which, while not a form of replication, provides an unbiased view of the evidence using the UK Biobank cohort. We identified two phenotypes detailing treatment by a psychiatrist or a doctor, respectively, for nerves, anxiety, tension or depression, but we did not find evidence of a causal effect of BMI on these phenotypes. A third result, the frequency of tenseness / restlessness in the last 2 weeks, was weakly associated in a direction consistent with the other nervousness associations we identified, although it did not pass the P value threshold corresponding to an estimated FDR of 5%.

Several previous observational studies have reported the association of anxiety and BMI. The prevalence of anxiety has been shown to be higher in obese compared with non-obese people ^{43–47}, although Vainik et al. found no correlation between BMI and neuroticism ⁴⁸. The few instrumental variable studies estimating the causal effect of BMI on anxiety that have been performed to date (one specifically looking at phobic anxiety ⁴⁹ and the other using an anxiety measure defined using the Hospital Anxiety and Depression Scale (HADS) ⁵⁰), did not find evidence of a causal effect, although this may be due to insufficient statistical power ^{49,50}. Two recent MR studies provided evidence that an increase in BMI adversely affects risk of depression symptoms ⁵¹ and major depressive disorder ⁵². Furthermore, it has been shown that the mechanism through which BMI-associated genetic variants affect risk of obesity may involve

regions of the brain such as the basal ganglia ⁵³, which are plausibly involved in emotional processes ⁵⁴.

The associations we identified with self-reported nervousness should be further investigated and replicated in an independent sample. These associations may reflect a true causal effect of BMI. Alternative explanations for these results include chance, or because some variants in the BMI genetic score have horizontal pleiotropic effects and are thus invalid instruments for BMI. We assessed the possibility that our instrument is invalid by comparing the estimates of two independent genetic instruments (i.e. *FTO* SNP versus the remaining genetic variants), and using alternative approaches that estimate the causal effects under differing assumptions of instrument validity – MR-Egger, weighted median and MBE. While we found little evidence of directional pleiotropy, we did find evidence of non-directional pleiotropy for the ‘being a worrier’ outcome, indicated by different estimates using the 95 SNP score versus *FTO* SNP, and also different estimates using IV probit regression compared to the weighted MBE.

UK Biobank is a highly-selected sample of the UK population, having a response rate of 5.5% ⁵⁵, that is not representative of the UK general population ⁵⁶. For example, UK Biobank participants have, on average, a lower BMI, and fewer self-reported health conditions, compared with the general population ⁵⁶. Hence, our estimates may be biased if selection into the sample is affected by BMI ⁵⁷. Also, if selection is additionally dependent on a given outcome (e.g. self-reported nervousness), associations may be biased by a particular form of collider bias – selection induced collider bias ⁵⁸. In general, collider bias may occur when two variables (A and B) independently affect a third variable (C) and variable C is conditioned upon in analyses. Selection induced

collider bias may occur when variable C represents whether a person is selected into the sample, i.e. variables A and B both independently affect participation in the study. Hence, estimates of association between two phenotypes – such as our BMI genetic score and a given outcome in our study – can be biased, if inclusion in the study is affected by both phenotypes.

We found that our BMI genetic score was associated with assessment centre, even after adjustment for the first 40 principal components. This may indicate that the genetic principal components are not fully accounting for genetic population differences. However, this may also be due to selection induced collider bias, because both BMI and location are related to selection into the sample⁵⁶. If both BMI and assessment centre affect participation in the study, then an association may be induced between the BMI genetic score and assessment centre. For example, the South West region had the highest participation rate of the regions sampled, such that living in this region is associated with a higher chance of participating compared to the other regions⁵⁶. Since BMI is negatively associated with participation in UK Biobank, we would expect (under most realistic assumptions about the association between BMI, location and participation in UK Biobank) to see a positive association between the BMI genetic risk score, and participating in the South West region compared to other regions if collider bias is the cause, and (using attendance at the Bristol assessment centre as a proxy for region of residence) this is indeed what we see (see illustration shown in Supplementary figure 8). Furthermore, the West Scotland region has the lowest participation rate, and attending the Glasgow assessment centre compared to other centres is associated with a lower BMI genetic risk score, which is expected if this relationship is due to selection induced collider bias.

We now discuss some further limitations of this work. PHESANT uses a rule-based method to automatically determine how to test each outcome, and it is possible that this may deal with some variables inappropriately. Also, PHESANT tests the linear association of the genetic score with the set of outcomes, and it is possible that non-linear associations may exist. For instance, it is possible that BMI has a non-linear effect on nervousness/anxiety (for example low and high BMI may cause higher levels of anxiety compared with those in the normal range) and in follow-up analyses this should be investigated. Ranking the associations means that we should expect the true strength of the associations to be less than we reported due to the winner's curse. We used stringent P value thresholds, which, although reducing the type I error rate, is likely to also increase the type II error rate. We used BMI as a surrogate measure for adiposity, but the effect of adiposity estimated using observational data is not the true effect of an intervention that modifies BMI, because this will depend on the particular intervention used. The causal effects we have estimated may be the result of changes of other aspects of BMI in addition to or instead of changes in adiposity (e.g. changes in lean mass versus fat mass)⁵⁹. Similarly, our BMI genetic score is an instrument for life-long BMI, hence we cannot say that it is BMI at a specific age (e.g. the age of the UK Biobank cohort) that affects an outcome. However, searching for potential causal effects using MR is useful to identify outcomes that may be modified through interventions on modifiable determinants of adiposity (e.g. diet or physical activity)⁵⁹.

PHESANT is the first tool to perform comprehensive phenome scans, where previously the set of outcomes tested would be restricted to a homogeneous subset³⁰. In this study, we have presented the first comprehensive phenome scan, using the PHESANT tool to

search for the causal effects of BMI. This MR-pheWAS confirmed several established effects of BMI, and also identified potentially interesting novel results, such as a potential causal effect of BMI on feelings of nervousness. This work demonstrates how MR-pheWAS can be applied in UK Biobank and can serve as a model for future studies. There is much potential to use MR-pheWAS to search for the causal effects of phenotypes where, compared to BMI, much weaker priors regarding their causal effects. Phenome-wide scans are a hypothesis generation approach and hence identified associations should be followed-up in an independent sample.

METHODS

Study population

UK Biobank sampled 503 325 men and women in the UK aged between 37–73 years (99.5% were between 40 and 69 years)³¹. This cohort includes a large and diverse range of data from blood, urine and saliva samples and health and lifestyle questionnaires²⁷.

Of the 487 406 participants with genetic data, we removed 373 with genetic sex different to reported sex, and 471 with sex chromosome aneuploidy (identified as putatively carrying sex chromosome configurations that are not either XX or XY). We found no outliers in heterozygosity and missing rates, which would indicate poor quality of the genotypes. We removed 78 309 participants not of white British ancestry³². We removed 73 277 participants who were identified as being related, having a kinship coefficient denoting a third degree (or closer) relatedness³². We removed 8

individuals with withdrawn consent, giving a sample of 334 968 participants. A participant flow diagram is given in Figure 1.

BMI allele score

We created an allele score from 96 genetic variants previously found to be associated with BMI, in a recent GWAS meta-analysis by the GIANT consortium³³. The score was calculated as a sum of the number of BMI-increasing alleles, weighted by the effect size as reported in the GIANT GWAS (reported as a SD change of BMI per dosage increase)³³ (see Supplementary table 1), such that a higher allele score corresponds to a higher BMI, and was standardized to have a mean of zero and SD of 1.

Outcomes

The Biobank data showcase allows researchers to identify variables based on the field type (<http://biobank.ctsu.ox.ac.uk/showcase/list.cgi>). At the time of initial data download there were 2143 fields of the following types: integer, continuous, categorical (single) and categorical (multiple) (identified in March 2016, and updated in November 2017 for fields where data accrual was still underway).

We excluded 70 fields a priori, given in Supplementary table 2, for the following reasons. We removed one field denoting the assessment centre. We removed 2 fields described by UK Biobank as ‘polymorphic’, containing values with mixed data types. We removed 14 fields that, although listed in the data showcase, were not currently available. We removed 13 genetic descriptor fields, 1 sex field and 4 age fields. We

removed 17 variables describing the assessment centre environment. We removed 18 categorical (single) fields with more than one value recorded per person.

We assigned 92 outcomes that denote adiposity or some aspect of weight, fat mass or height as ‘exposure phenotypes’ – to be included in our MR-pheWAS analysis so that we can assess the strength of outcome associations in relation to these, while this a priori assignment means they will not contribute to the multiple testing burden (see Supplementary table 3).

This resulted in a set of 1981 UK Biobank fields (263 integer, 969 continuous, 649 categorical (single) and 100 categorical (multiple)), referred to hereafter as the outcome dataset (because they are tested as an outcome irrespective of whether this is biologically plausible).

Observed BMI

Weight and height were measured at the initial UK Biobank assessment centre – weight in light clothing and unshod was measured using Tanita BC418MA body composition analyser to the nearest 100g, and height to the nearest cm using a Seca 202 device. These were used to calculate BMI (kg/m^2).

Covariates

We include age and sex as covariates in our models to reduce the variation in our outcomes. Age when participants attended the UK Biobank assessment centre was

derived from their date of birth and the date of their assessment centre visit. Sex was self-reported during the touchscreen questionnaire (and validated using the genome-wide data). We adjust for the first 10 genetic principal components to control for confounding via population stratification. Genetic variants are set at conception, and after conception they cannot be affected by traditional confounding factors, therefore we did not adjust for any further covariates.

Statistical methods

PHESANT MR-pheWAS

We test the direct association of the BMI genetic score with each of the outcome variables using the PHESANT package (version 0.10). A description of PHESANT's automated rule-based method is given in detail elsewhere³⁰. In brief, the decision rules start with the variable field type and use rules to categorize each variable as one of four data types: continuous, ordered categorical, unordered categorical or binary. Variables with the continuous and integer field type are usually assigned to the continuous data type, but some are assigned to ordered categorical if, for instance, there are only a few distinct values. Variables of the categorical (single) field type are assigned to either the binary, ordered categorical or unordered categorical, depending on whether the field has two distinct values, or has been specified as ordered or unordered in the PHESANT setup files. Variables of the categorical (multiple) field type are converted to a set of binary variables, one for each value in the categorical (multiple) fields.

PHESANT estimates the bivariate association of the BMI genetic score with each outcome variable. The BMI genetic score and outcome variables are the independent

(exposure) and dependent (outcome) variables in the regression model, respectively. Outcome variables with continuous, binary, ordered categorical and unordered categorical data types, are tested using linear, logistic, ordered logistic, and multinomial logistic regression, respectively. Prior to testing, an inverse normal rank transform is applied to variables of the continuous data type, to ensure they are normally distributed. All analyses are adjusted for covariates as described above.

We correct for multiple testing by controlling for the expected proportion of false positive results. After ranking the results by P value, we identify the largest rank position with a P value less than $P_{threshold} = 0.05 \times \text{rank}/n$, where n is the total number of tests in the phenome scan. $P_{threshold}$ is the P value threshold resulting in a false discovery rate of 5%³⁴. We also calculate a stringent Bonferroni corrected P-value threshold, by dividing 0.05 by the number of tests performed, which assumes each test is independent.

Results visualization with PHESANT-viz

We use PHESANT-viz, a D3 Javascript visualization tool included in the PHESANT package, to visualize our results as a graph, using the Biobank assigned category structure (<http://biobank.ctsu.ox.ac.uk/showcase/label.cgi>). This enables interpretation of the identified associations with consideration of the results for related variables. For example, the estimated effects of BMI on depression are grouped with the results of other psychosocial phenotypes.

PHESANT sensitivity analysis

We re-run our phenome scan to assess residual confounding, additionally adjusting for both assessment centre and genetic batch.

Testing for residual population stratification

We test the extent that genetic principal components account for genetic differences across the population. We test the association of assessment centre with the BMI genetic score (as the independent and dependent variables, respectively), adjusting for:

- 1) Age and sex.
- 2) Age, sex and the first 10 genetic principal components.
- 3) Age, sex and the first 40 genetic principal components.

We use a likelihood ratio test to determine the strength of the association of the assessment centres collectively, with the genetic score. An association between the assessment centres and the BMI genetic score, after adjusting for the genetic principal components, may indicate that the genetic principal components are not fully accounting for population stratification.

Follow-up analysis of identified associations

We identified associations with a related set of psychosocial traits and generate a QQ-plot restricting to the psychosocial UK Biobank category only (category ID=100059), to determine whether these results have an association stronger than expected by chance, given the results of related phenotypes. We perform a formal instrumental variable analysis for a set of binary outcomes, using two-stage IV probit regression (the Stata `ivprobit` command, with conditional maximum-likelihood estimation), adjusting for age, sex and the first 10 genetic principal components. We also adjust for

age, sex and the first 40 genetic principal components, as a sensitivity analysis. We take the exponent of 1.6 times the estimates, to approximate the association in terms of the change of odds ³⁵. We also test the observational association of BMI with each phenotype using logistic regression (Stata `logistic` command).

We compare estimates within the initial discovery sample (used in our initial presentation of PHESANT ³⁰), with estimates using the additional participants used in this study, as a replication. We remove all participants from our replication sample who were identified as being related to a participant in the discovery sample, having a kinship coefficient denoting a third degree (or closer) relatedness.

Follow-up sensitivity analyses

We performed sensitivity analyses to investigate whether our genetic score, and its constituent genetic variants, may be invalid instruments for BMI. First, we compared the effect estimates using two independent instrumental variables: 1) SNP rs1558902 (the SNP most strongly associated with BMI, at the *FTO* locus), and 2) the remaining 95 genetic variants.

Second, we also estimated causal effects using three alternative MR approaches – MR-Egger ³⁶, weighted median ³⁷ and MBE ³⁸. These methods estimate effects consistent with the true causal effect under more relaxed assumptions of instrument validity, compared with IV probit regression where for example, including just one genetic variant with pleiotropic effects could bias estimates. These methods are performed using two sample MR, using estimates of association of each genetic variant on BMI, and estimates of association of each genetic variant on a given outcome, respectively,

estimated in different populations. We use the SNP-BMI estimates from the GIANT GWAS³³ (that did not include UK Biobank), and estimate the SNP-outcome associations on our UK Biobank sample, adjusting for age, sex and the first 10 genetic principal components.

The MR-Egger, weighted median and MBE approaches are complementary, each depending on distinct assumptions about instrument validity. Estimates from MR-Egger are not biased by horizontal pleiotropy, where a genetic variant affects several traits through separate pathways, under the assumption that the association of each genetic variant with the exposure is independent of any horizontal pleiotropic effect – referred to as the InSIDE (Instrument Strength Independent of Direct Effects) assumption³⁶. This approach can test for directional pleiotropy, which is when the horizontal pleiotropic effects of the genetic variants are not balanced about the null³⁶. Directional pleiotropy is identified where the intercept estimate is not consistent with the null. MR-Egger assumes that the SNP-exposure associations are measured without error, known as the NOME assumption³⁹. We determine the degree to which the NOME assumption is violated, using the I_{GX}^2 statistic, which is an adaption of the I^2 statistic used in the field of meta-analysis³⁹. The I_{GX}^2 statistic ranges between 0 and 1, and captures the uncertainty in the estimated SNP-exposure associations relative to the heterogeneity across the true underlying SNP-exposure associations. It is recommended that results of MR-Egger should be treated with caution when $I_{GX}^2 < 0.9$ ³⁶. Furthermore, we performed an MR-Egger analysis with bias adjustment using SIMEX, to estimate the causal effect when the NOME assumption is violated³⁹. In brief, SIMEX works by learning models with increasing violations of the NOME assumption (i.e. increasing error in the SNP-exposure associations). The estimates from these models

are then treated as a set of data points and a model is learnt across these. This ‘meta’ model is then used to extrapolate back to the estimate that would have occurred if the NOME assumption was satisfied.

The weighted median approach estimates a consistent causal effect under the assumption that less than 50% of the genetic variants are invalid instruments³⁷. The simple version of MBE assumes that the mode of the smoothed empirical density function of SNP estimates is consistent with the true causal effect, even if the majority of the SNP estimates are not consistent – known as the ZERo Mode Pleiotropy Assumption (ZEMPA)³⁸. We also test using a weighted version of the MBE, which instead assumes that the mode of the inverse-variance weighted empirical density function is consistent with the true causal effect. The MBE method uses a bandwidth parameter ϕ , which determines the amount of smoothing of the empirical density. We generate the smoothed empirical density using a range of ϕ values, and choose a ϕ value that provides an appropriate amount of smoothing (a degree of smoothing that allows multi-modal distributions to be identified while not overfitting to the estimated values). We test the simple and weighted versions of the MBE using our chosen ϕ value.

Analyses are performed in R version 3.2.4 ATLAS, Matlab r2015a or Stata version 14, and code is available at [<https://github.com/MRCIEU/PHESANT-MR-pheWAS-BMI>]. Git tag v0.2 corresponds to the version presented here.

REFERENCES

1. Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**, 569–578 (2008).
2. Must, A. *et al.* The disease burden associated with overweight and obesity. *JAMA* **282**, 1523–1529 (1999).
3. Visscher, T. L. S. & Seidell, J. C. The Public Health Impact of Obesity. *Annu. Rev. Public Health* **22**, 355–375 (2001).
4. Davey Smith, G. & Ebrahim, S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
5. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
6. Richmond, R. C. *et al.* Assessing causality in the association between child adiposity and physical activity levels: a Mendelian randomization analysis. *PLoS Med.* **11**, e1001618 (2014).
7. Würtz, P. *et al.* Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med.* **11**, e1001765 (2014).
8. Timpson, N. J., Sayers, A., Davey-Smith, G. & Tobias, J. H. How Does Body Fat Influence Bone Mass in Childhood? A Mendelian Randomization Approach. *J. Bone Miner. Res.* **24**, 522–533 (2009).
9. Mumby, H. S. *et al.* Mendelian randomisation study of childhood BMI and early menarche. *J. Obes.* **2011**, 180729 (2011).
10. Brennan, P. *et al.* Obesity and cancer: Mendelian randomization approach utilizing the FTO genotype. *Int. J. Epidemiol.* **38**, 971–975 (2009).
11. Kivimäki, M. *et al.* Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young. *Eur. Heart J.* **29**, 2552–60 (2008).
12. Welsh, P. *et al.* Unraveling the directional link between adiposity and inflammation: a bidirectional Mendelian randomization approach. *J.* **95**, 93–99 (2010).
13. Hubacek, J. A. *et al.* The FTO gene polymorphism is associated with end-stage renal disease: two large independent case–control studies in a general population. *Nephrol. Dial. Transplant.* **27**, 1030–1035 (2011).
14. Freathy, R. M. *et al.* Common Variation in the FTO Gene Alters Diabetes-Related Metabolic Traits to the Extent Expected Given Its Effect on BMI. *Diabetes* **57**, 1419–1426 (2008).
15. Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–94 (2007).
16. Nordestgaard, B. G. *et al.* The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian Randomisation Approach. *PLoS Med.* **9**, e1001212 (2012).
17. Noyce, A. J. *et al.* Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Med.* **14**, e1002314 (2017).

18. Benn, M., Tybjaerg-Hansen, A., Davey Smith, G. & Nordestgaard, B. G. High body mass index and cancer risk—a Mendelian randomisation study. *Eur. J. Epidemiol.* **31**, 879–892 (2016).
19. Corbin, L. J. *et al.* Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. *Diabetes* **65**, 3002–3007 (2016).
20. Tyrrell, J. *et al.* Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ* **352**, i582 (2016).
21. Skaaby, T. *et al.* Estimating the causal effect of body mass index on hay fever, asthma, and lung function using Mendelian randomization. *Allergy* (2017). doi:10.1111/all.13242
22. Lyall, D. M. *et al.* Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank: A Mendelian Randomization Study. *JAMA Cardiol.* **2**, 882–889 (2017).
23. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
24. Denny, J. C., Bastarache, L. & Roden, D. M. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annu. Rev. Genomics Hum. Genet.* **17**, 353–373 (2016).
25. Millard, L. A. C. *et al.* MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Sci. Rep.* **5**, 16645 (2015).
26. Cronin, R. M. *et al.* Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front. Genet.* **5**, 250 (2014).
27. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, e1001779 (2015).
28. Lyall, D. M. *et al.* Body-mass index and cardiometabolic disease: a Mendelian randomisation study of UK Biobank participants. *Lancet* **388**, S9 (2016).
29. Klarin, D., Emdin, C. A., Natarajan, P., Conrad, M. F. & Kathiresan, S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. *Circ. Cardiovasc. Genet.* **10**, (2017).
30. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* (2017).
31. Allen, N. *et al.* UK Biobank: Current status and what it means for epidemiology. *Heal. Policy Technol.* **1**, 123–126 (2012).
32. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017).
33. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
34. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
35. Rassen, J. A., Schneeweiss, S., Glynn, R. J., Mittleman, M. A. & Brookhart, M. A. Instrumental Variable Analysis for Estimation of Treatment Effects With Dichotomous Outcomes. *Am. J. Epidemiol.* **169**, 273–284 (2008).
36. Bowden, J., Smith, G. & Burgess, S. Mendelian randomization with invalid

- instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
37. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
38. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **28**, 30–42 (2017).
39. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I(2) statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
40. Timpson, N. J. *et al.* Does greater adiposity increase blood pressure and hypertension risk? Mendelian randomization using the FTO/MC4R genotype. *Hypertension* **54**, 84–90 (2009).
41. White, I. R., Daniel, R. & Royston, P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables(). *Comput. Stat. Data Anal.* **54**, 2267–2275 (2010).
42. Sterne, J. A. C. & Davey Smith, G. Sifting the evidence—what’s wrong with significance tests? Another comment on the role of statistical methods. *BMJ* **322**, 226–231 (2001).
43. Garipey, G., Nitka, D. & Schmitz, N. The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. *Int. J. Obes.* **34**, 407–419 (2010).
44. Bodenlos, J. S., Lemon, S. C., Schneider, K. L., August, M. A. & Pagoto, S. . Associations of mood and anxiety disorders with obesity: Comparisons by ethnicity. *J. Psychosom. Res.* **71**, 319–324 (2011).
45. Strine, T. W. *et al.* Depression and Anxiety in the United States: Findings From the 2006 Behavioral Risk Factor Surveillance System. *Psychiatr. Serv.* **59**, 1383–1390 (2008).
46. Bjerkeset, O., Romundstad, P., Evans, J. & Gunnell, D. Association of Adult Body Mass Index and Height with Anxiety, Depression, and Suicide in the General Population: The HUNT Study. *Am. J. Epidemiol.* **167**, 193–202 (2007).
47. Zhao, G. *et al.* Depression and anxiety among US adults: associations with body mass index. *Int. J. Obes.* **33**, 257–266 (2009).
48. Vainik, U. *et al.* Neurobehavioural Correlates of Obesity are Largely Heritable. *bioRxiv* (2017). doi:10.1101/204917
49. Walter, S. *et al.* Do genetic risk scores for body mass index predict risk of phobic anxiety? Evidence for a shared genetic risk factor. *Psychol. Med.* **45**, 181–191 (2015).
50. Bjørngaard, J. H. *et al.* Association of Body Mass Index with Depression, Anxiety and Suicide—An Instrumental Variable Analysis of the HUNT Study. *PLoS One* **10**, e0131708 (2015).
51. van den Broek, N. *et al.* Causal Associations Between Body Mass Index and Mental Health: A Mendelian Randomization Study. *bioRxiv* (2017).
52. Wray, N. R. & Sullivan, P. F. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv* (2017).
53. Ghosh, S. & Bouchard, C. Convergence between biological, behavioural and genetic determinants of obesity. (2017).

54. Lanciego, J. L., Luquin, N. & Obeso, J. A. Functional Neuroanatomy of the Basal Ganglia. *Cold Spring Harb. Perspect. Med.* **2**, a009621 (2012).
55. Swanson, J. M. The UK Biobank and selection bias. *Lancet* **380**, 110 (2012).
56. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* (2017). doi:10.1093/aje/kwx246
57. Hughes, R. A., Davies, N. M., Davey Smith, G. & Tilling, K. Selection bias in instrumental variable analyses. *bioRxiv* (2017). doi:10.1101/192237
58. Munafo, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider Scope: When selection bias can substantially influence observed associations. *Int. J. Epidemiol.* dx206 (2017). doi:<https://doi.org/10.1101/079707>
59. Hernan, M. A. & Taubman, S. L. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes* **32**, S8–S14

ACKNOWLEDGEMENTS

This work was supported by the University of Bristol and UK Medical Research Council [grant numbers MC_UU_12013/1, MC_UU_12013/8 and MC_UU_12013/9]. LACM is funded by a University of Bristol Vice-Chancellor's Fellowship. This research has been conducted using the UK Biobank Resource under Application Number 16729.

AUTHOR CONTRIBUTIONS

LACM contributed to the design of the study, performed all analyses, wrote the first version of the manuscript, critically reviewed and revised the manuscript and approved the final version of the manuscript as submitted. NMD contributed to the design of the study, critically reviewed and revised the manuscript and approved the final version of the manuscript as submitted. KT contributed to the design of the study, critically reviewed and revised the manuscript and approved the final version of the manuscript as submitted. TRG critically reviewed and revised the manuscript and approved the final version of the manuscript as submitted. GDS conceptualized the study, contributed to the design of the study, critically reviewed and revised the manuscript and approved the final version of the manuscript as submitted.

COMPETING FINANCIAL INTERESTS STATEMENT

The authors declare no competing financial interests.

FIGURES

Figure 1: Participant flow diagram

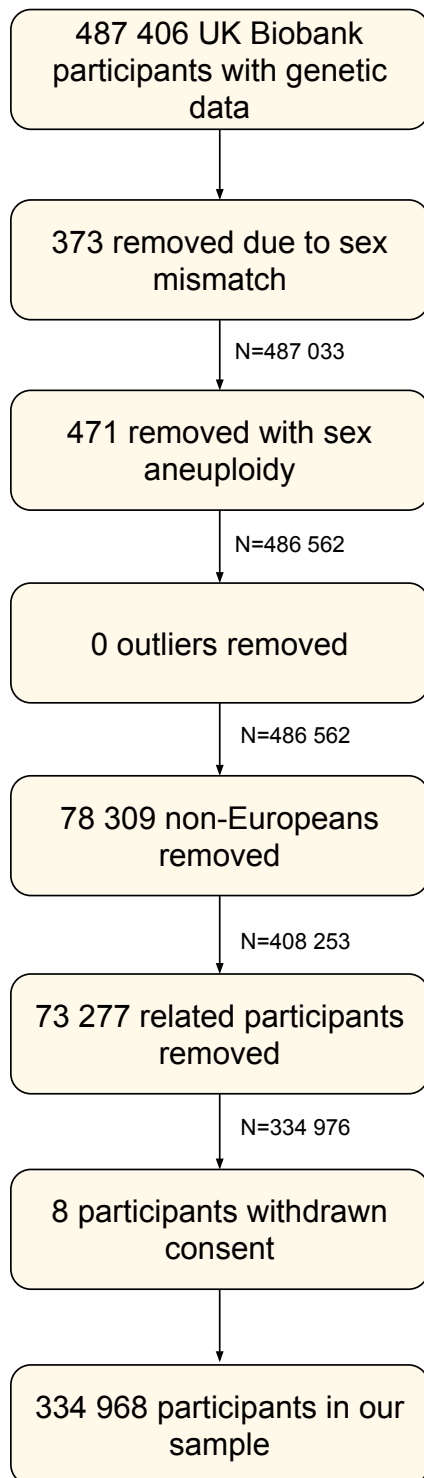
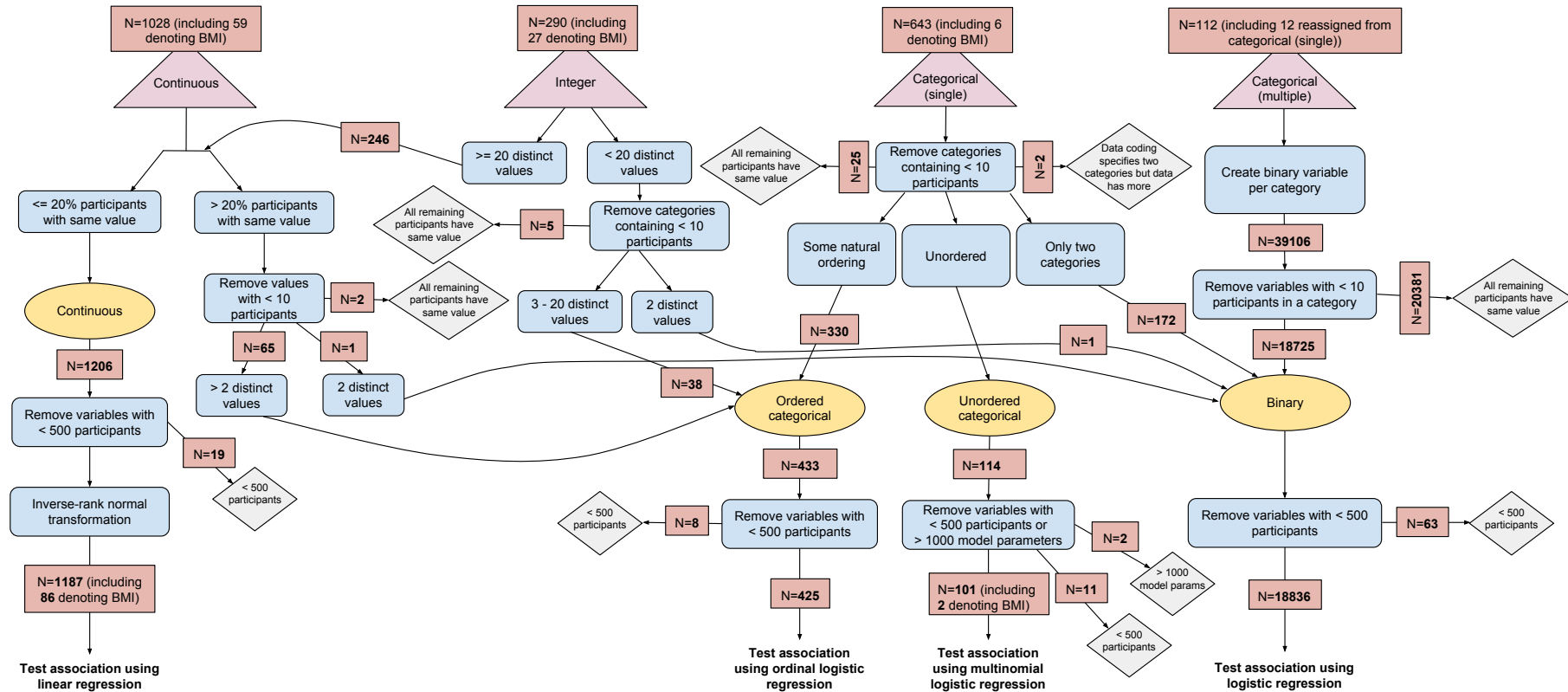
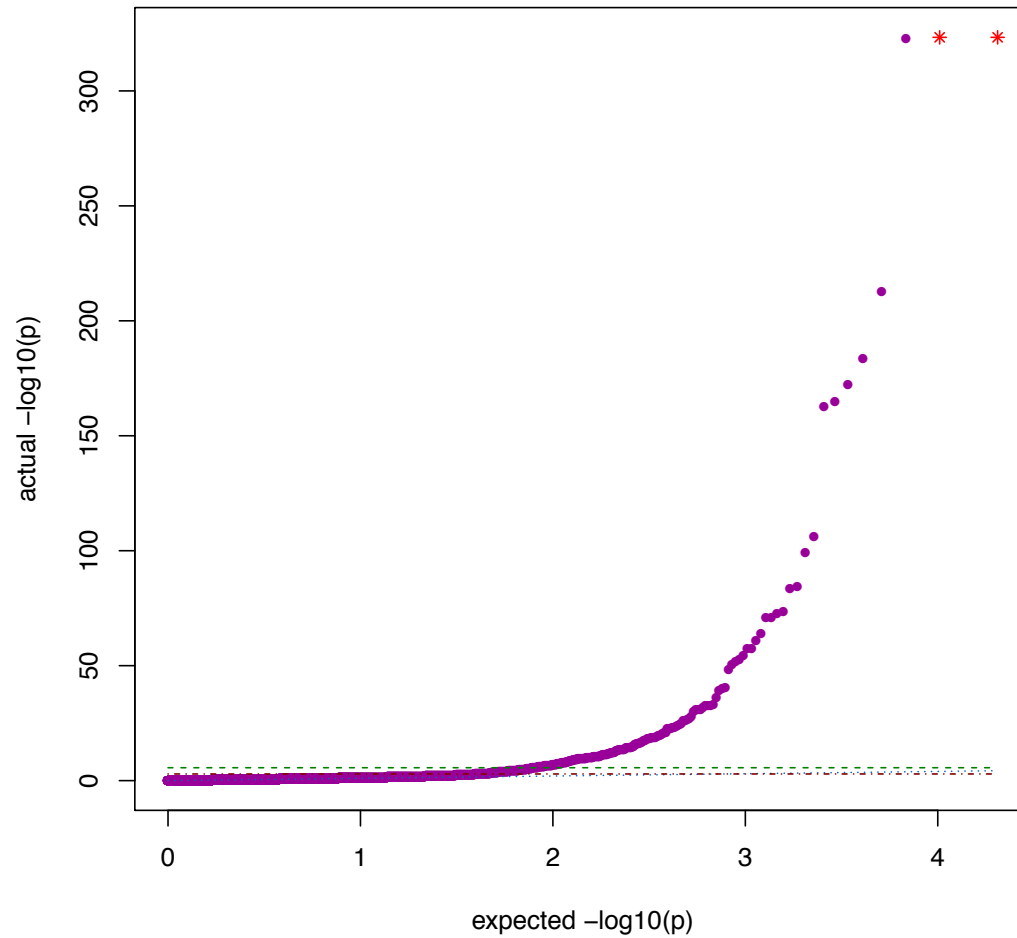


Figure 2: Variable processing flow diagram showing logic from defined field type in UK Biobank data to test of association, with number of variables reaching each stage of processing flow



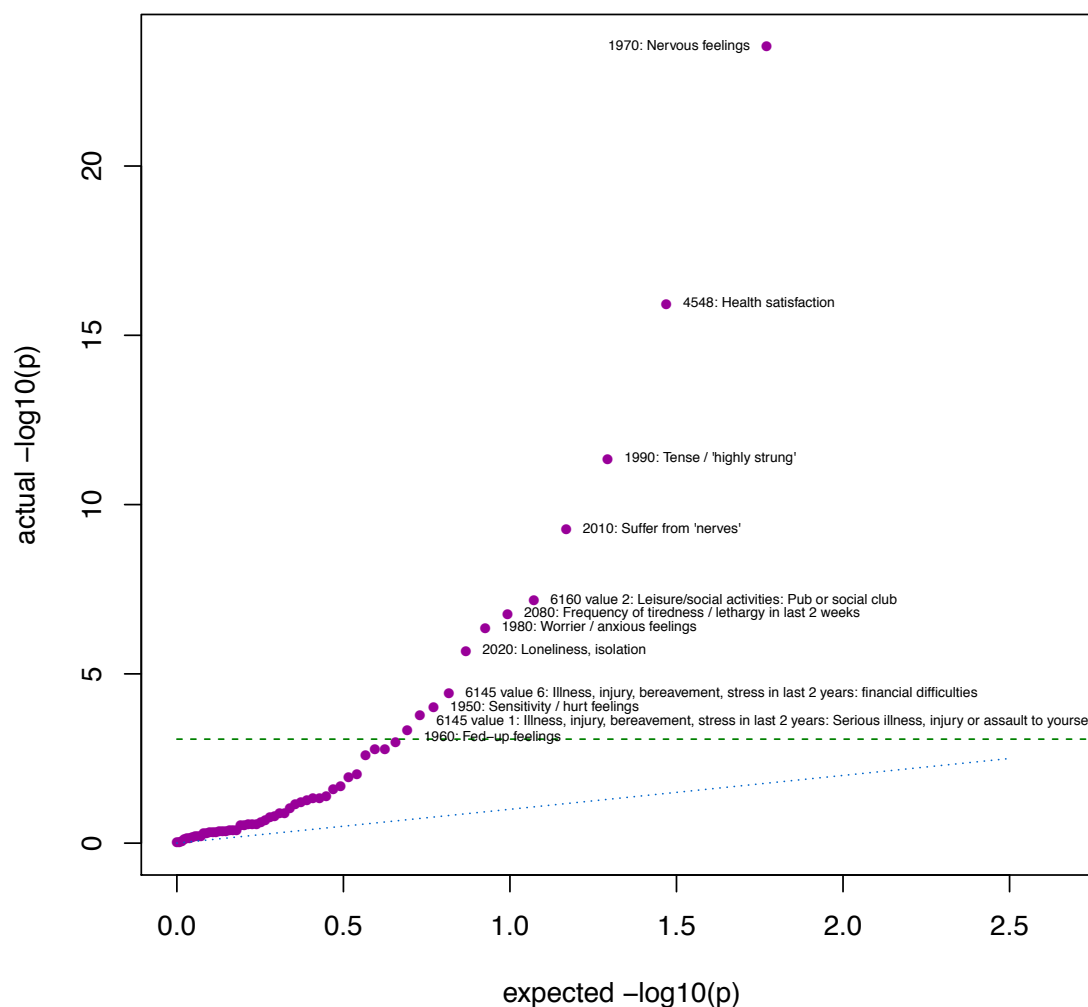
Variable processing flow diagram showing logic from defined field type specified by UK Biobank to test of association, and the number of variables at each stage of the processing flow. Triangular nodes at top of figure are field types defined by UK Biobank. Rectangular nodes show processing logic used to determine the data type assignment (oval), either continuous, ordered categorical, unordered categorical or binary, and hence finally, the type of test used: linear, ordinal logistic, multinomial logistic or logistic regression, respectively. Diamond nodes show points where variables may be removed.

Figure 3: QQ plot of 20,461 MR-pheWAS results



Green dashed line: Bonferroni corrected threshold ($p = 2.44 \times 10^{-6}$). Red dash-dotted line: FDR threshold ($p = 1.27 \times 10^{-3}$). Blue dotted line: actual = expected. Purple points: results of tests performed in MR-pheWAS. Red stars: results with P values $< 9.88 \times 10^{-324}$.

Figure 4: QQ plot of results in UK Biobank field category ‘psychosocial factors’ (ID 100059)



Green dashed line: Bonferroni corrected threshold calculated for psychosocial traits only ($0.05/59 = 8.47 \times 10^{-4}$). Blue dotted line: actual = expected.

Purple points: results of tests performed in MR-pheWAS.

TABLES

Table 1: Results of follow-up analysis of nervousness / worrying traits

	Nervous person (FID=1970)	Being a worrier (FID=1980)	Tense / highly strung (FID=1990)	Suffering from nerves (FID=2010)
N	325,552	325,523	324,146	322,236
OBSERVATIONAL ESTIMATES⁷				
Observational association ^{3*}	0.964 [0.962, 0.966]	0.981 [0.980, 0.983]	0.983 [0.981, 0.985]	0.993 [0.991, 0.994]
CAUSAL ESTIMATES				
Estimates using two-stage IV probit regression⁸				
Effect estimate with 96 SNP score ¹	-0.039 [-0.047, -0.032]	-0.018 [-0.025, -0.011]	-0.028 [-0.036, -0.020]	-0.024 [-0.032, -0.017]
Effect estimate with 96 SNP score – approximate change of odds ^{4*}	0.939 [0.928, 0.950]	0.972 [0.961, 0.983]	0.956 [0.944, 0.968]	0.962 [0.950, 0.974]
Effect estimate with 96 SNP score, in terms of SD change of BMI ⁶	-0.186 [-0.221, -0.151]	-0.085 [-0.117, -0.052]	-0.134 [-0.172, -0.096]	-0.115 [-0.152, -0.079]
Effect estimate with 96 SNP score, in terms of SD change of BMI, approximate change of log odds ⁶⁺	-0.297 [-0.354, -0.241]	-0.135 [-0.188, -0.083]	-0.214 [-0.275, -0.154]	-0.185 [-0.243, -0.127]
Effect estimate with 95 SNP score (excluding FTO SNP) ¹	-0.040 [-0.048, -0.032]	-0.022 [-0.030, -0.015]	-0.032 [-0.040, -0.023]	-0.027 [-0.035, -0.019]
Effect estimate with FTO SNP ¹	-0.031 [-0.050, -0.012]	0.012 [-0.006, 0.029]	-0.004 [-0.025, 0.017]	-0.008 [-0.027, 0.012]
Estimates using MR-Egger²				
Causal effect estimate ⁺	-0.183 [-0.405, 0.040]	-0.036 [-0.223, 0.150]	-0.123 [-0.364, 0.119]	-0.170 [-0.400, 0.060]
Intercept estimate	-0.002 [-0.008, 0.004]	-0.002 [-0.007, 0.003]	-0.002 [-0.009, 0.004]	0.000 [-0.006, 0.007]
I^2_{GX} statistic	0.893			
Bias adjusted causal effect estimate ⁵⁺	-0.204 [-0.452, 0.043]	-0.039 [-0.248, 0.170]	-0.135 [-0.405, 0.135]	-0.193 [-0.449, 0.064]
Estimates using weighted median²				
Weighted median effect estimate ⁺	-0.220 [-0.315, -0.125]	-0.087 [-0.166, -0.009]	-0.172 [-0.269, -0.075]	-0.087 [-0.180, 0.007]
Estimates using mode based estimate²				
Simple, $\phi=0.75^+$	-0.086 [-0.317, 0.146]	-0.161 [-0.453, 0.131]	-0.273 [-0.600, 0.055]	-0.019 [-0.303, 0.266]

Weighted, $\phi=0.75^+$	-0.158 [-0.286, -0.031]	0.063 [-0.053, 0.180]	-0.114 [-0.272, 0.043]	-0.048 [-0.182, 0.085]
DISCOVERY AND REPLICATION ESTIMATES^{9,10}				
Discovery effect estimate with 96 SNP score ¹	-0.041 [-0.055, -0.027]	-0.022, [-0.035, -0.008]	-0.035 [-0.051, -0.020]	-0.029 [-0.044, -0.014]
Discovery Effect estimate with 96 SNP score – approximate change of odds ^{4*}	0.936 [0.915, 0.958]	0.966 [0.946, 0.987]	0.945 [0.922, 0.968]	0.955 [0.933, 0.978]
Replication effect estimate with 96 SNP score ¹	-0.040 [-0.049, -0.031]	-0.015 [-0.023, -0.006]	-0.024 [-0.034, -0.014]	-0.023 [-0.033, -0.014]
Replication Effect estimate with 96 SNP score – approximate change of odds ^{4*}	0.938 [0.925, 0.952]	0.977 [0.964, 0.990]	0.963 [0.948, 0.978]	0.963 [0.949, 0.978]

BMI: body mass index; FID: field identifier; SD: standard deviation; IV: instrumental variable.

¹ Coefficient [95% confidence interval] from instrument variable probit models (ivprobit Stata command), for a 1 kg/m² higher BMI.

² Two sample analyses use the SNP-BMI associations from Locke GWAS³³, and SNP-outcome associations estimated in UK Biobank (adjusted for and first 10 genetic principal components).

³ Estimate of the odds ratio of outcome for a 1 kg/m² increase in BMI.

⁴ Estimate of the odds ratio of outcome for a 1 kg/m² increase in BMI, calculated by taking the exponent of 1.6 times the probit estimate³⁵.

⁵ Bias adjusted causal effect estimate using MR-Egger with SIMEX.

⁶ Effect estimate for SD change in BMI and logistic approximation, for comparison with two-sample approaches, as these use the BMI standardised SNP associations from the GIANT GWAS³³.

⁷ Adjusted for age and sex.

⁸ Adjusted for age, sex and first 10 genetic principal components (see Supplementary table 6 for results of sensitivity analysis, adjusting for age, sex and first 40 genetic principal components).

⁹ Discovery results are estimated on the UK Biobank sample used in the PHESANT application note usage example³⁰.

¹⁰ Sample size for discovery samples: Field 1970: 83,843; Field 1980: 83,795; Field 1990: 83,445; Field 2010: 82,925.

Sample size for replication samples: Field 1970: 216,311; Field 1980: 216,301; Field 1990: 215,435; Field 2010: 214,152.

⁺ Estimates are in terms of change of log odds of outcome variable for a 1 SD higher BMI.

^{*} Estimates are in terms of odds ratio of outcome variable for a 1 kg/m² higher BMI.