Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration

Journal:	*BMJ*
Manuscript ID	Draft
Article Type:	Research methods and reporting
BMJ Journal:	BMJ
Date Submitted by the Author:	n/a
Complete List of Authors:	Skrivankova, Veronika; University of Bern, Institute of Social and Preventive Medicine Richmond, Rebecca; University of Bristol, Medical Research Council Integrative Epidemiology Unit; University of Bristol, Population Health Sciences, Bristol Medical School Woolf, Benjamin; University of Bristol, Medical Research Council Integrative Epidemiology Unit; University of Bristol, Department of Psychological Science Davies, Neil; University of Bristol, MRC Integrative Epidemiology Unit (IEU); University of Bristol, Population Health Sciences Swanson, Sonja; Erasmus MC, Department of Epidemiology VanderWeele, Tyler ; Harvard University T H Chan School of Public Health, Department of Epidemiology Timpson, Nicholas; University of Bristol, Medical Research Council Integrative Epidemiology Unit; University of Bristol, Population Health Sciences Higgins, Julian; University of Bristol, Population Health Sciences; NIHR Bristol Biomedical Research Centre Dimou, Niki; International Agency for Research on Cancer, Nutrition and Metabolism Branch Langenberg, Claudia; University of Cambridge, MRC Epidemiology Unit; Charité Universitätsmedizin Berlin, Berlin Institute of Health Loder, Elizabeth; Brigham and Women's Hospital, Division of Headache, Department of Neurology; Harvard Medical School Golub, Robert; Northwestern University Feinberg School of Medicine, Department of Medicine; JAMA Egger, Matthias; University of Bern, Institute of Social and Preventive Medicine; University of Bristol, Population Health Sciences, Bristol Medical School Davey Smith, George; University of Bristol, Medical Research Council Integrative Epidemiology Unit; University of Bristol, Population Health Sciences, Bristol Medical School Richards, Brent; McGill University, Departments of Medicine, Human Genetics, Epidemiology & Biostatistics, Lady Davis Institute, Jewish General Hospital; University of London, Department of Twin Research and Genetic Epidemiology, King's College London

Keywords:	Mendelian randomisation, Reporting guidelines, Checklist, Transparent reporting, Explanation and elaboration

Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration

Veronika W Skrivankova[1], Rebecca C Richmond[2,3], Benjamin AR Woolf[2,4], Neil M Davies[2,3,5], Sonja A Swanson[6], Tyler J VanderWeele[7], Nicholas J Timpson[2,3], Julian PT Higgins[3,8], Niki Dimou[9], Claudia Langenberg[10,11], Elizabeth W Loder[12,13], Robert M Golub[14], Matthias Egger[1,3,15], George Davey Smith[2,3],* J Brent Richards[16,17]**

Affiliations:

[1] *Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland*

[2] *Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom*

[3] *Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom*

[4] *Department of Psychological Science, University of Bristol, Bristol, United Kingdom*

[5] *K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.*

[6] *Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands*

[7] *Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America*

[8] *NIHR Bristol Biomedical Research Centre, Bristol, United Kingdom*

[9] *Nutrition and Metabolism Branch, International Agency for Research on Cancer (IARC/WHO), Lyon, France*

[10] *MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom*

[11] *Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Germany*

[12] *Division of Headache, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America*

[13] *Harvard Medical School, Boston, Massachusetts, United States of America*

[14] *Deputy Editor, JAMA, Chicago, Illinois; Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois;, United States of America*

[15] *Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa*

[16] *Departments of Medicine, Human Genetics, Epidemiology & Biostatistics, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Canada*

[17] *Department of Twin Research and Genetic Epidemiology, King's College London, University of London, London, United Kingdom*

Corresponding Author:

Prof Brent Richards
3755 Côte Ste-Catherine Road, Suite H-413 Montréal, QC H3T 1E2, Canada

Contact us: *contact@strobe-mr.org*
See also: *www.strobe-mr.org*

---

* Denotes equal contribution to this manuscript.

Word count: Standfirst 121 words; summary points 121 words; main text including boxes and tables 17,286 words; three figures (all from examples); 125 references; web appendix with nine examples, one figure and 1721 words.

Standfirst

Mendelian randomisation (MR) studies allow a better understanding of the causal effects of modifiable exposures on health outcomes, but the published evidence is often hampered by inadequate reporting. Reporting guidelines help authors effectively communicate all critical information about what was done and what was found. We developed STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation), to assist authors in reporting their MR research clearly and transparently. Adopting STROBE-MR should help readers, reviewers, and journal editors evaluate the quality of published MR studies. This article explains the 20 items of the STROBE-MR checklist, along with their meaning and rationale, using terms defined in a glossary. Examples of transparent reporting are used for each item to illustrate best practices.

Key messages

- In observational epidemiology, Mendelian randomisation (MR) studies provide an opportunity to study the causal relationship between an exposure and an outcome while reducing the risk of certain biases.
- There is little consensus around the reporting of MR studies, and the quality of reporting of these studies has been inconsistent. Many MR study reports do not state or examine the various assumptions of MR and report insufficient details on the data sources.
- We developed STROBE-MR, a checklist of 20 reporting items that should be considered when communicating MR studies. We explain their rationale and provide examples of transparent reporting.
- MR study authors, reviewers and journal editors are encouraged to use STROBE-MR to enable the potential of these studies to be fully realised.

Introduction

Observational epidemiology often examines the relationships between exposures and health outcomes. However, exposure-outcome associations reported in epidemiological studies are often not reliable estimates of causal effects. They can be produced by confounding, i.e., by another factor that affects both the outcome and exposure,[1 2 3] or other forms of bias. For example, alcohol consumption may be related to many potential confounding factors, including smoking, an unhealthy diet, and limited exercise. In turn, ill health may be related to a reduction or cessation of alcohol consumption, introducing potential bias due to reverse causality, when interest is in studying the effect of alcohol consumption on subsequent health.[4 5] There are several approaches to deal with such biases.[6] Instrumental variable methods rely on an external factor that determines the exposure of interest but is not associated with the outcome other than through its effect on the exposure.[6 7]

Over the past decade, advances in genetic technologies have enabled the identification of thousands of reproducible associations between genetic variation and relevant exposures, traits, and health outcomes. These genetic variations can be used as instrumental variables to shed light on the effect of modifiable exposures on diseases through a study method termed Mendelian randomisation (MR).[8] MR studies use genetic variants robustly related to modifiable exposures to understand the influence of the exposure on various health, social and economic outcomes. Genetic variation is essentially randomly inherited from parents to offspring at conception, and consequently, many factors that confound the relationship between the exposure and outcome cannot affect the genetic variants. Similarly, genetic variants are generally not influenced by the outcome and therefore, by reverse causation. This “Mendelian randomisation” thus provides an opportunity to study the relationship between exposures and outcomes while reducing potential bias from confounding and reverse causation.[9] These features make genetic variants suitable candidates for instrumental variable analysis, which can help estimate the causal effects of modifiable exposures on outcomes.[7] For example, the rs1229984 variant in the alcohol dehydrogenase 1B gene (*ADH1B*) has been used as an instrument to investigate the causal role of alcohol in cardiovascular disease.[10] Given these advantages, MR studies have increased in popularity and have begun to inform understanding of disease aetiology. As discussed in **Box 1**, MR is not limited to studies using genetic variants to generate instrumental variable estimates; however, these studies dominate the literature. A glossary of terms commonly used in MR is given in **Box 2**. Additional terms and explanations can be found in a comprehensive open-access MR dictionary.[11]

Box 1: The scope of Mendelian randomisation and STROBE-MR

While Mendelian randomisation (MR) generally uses genetic variation as the instrumental variable, MR is not limited to such studies. Indeed, the term "Mendelian randomisation" was introduced in 1991 for investigations of bone marrow transplantation in the treatment of childhood malignancies.[12 13] The basic notion was that if a child had an HLA compatible sibling, he or she was more likely to receive a bone marrow transplant than if there was not such a sibling. Analysing outcomes according to whether the child did or did not have such a sibling (optimally taking the number of siblings into account) is analogous to an intention to treat analysis in an RCT.[12 13] Having an HLA compatible sibling (which is a matter of chance) may also serve as a genetic 'instrument' for bone marrow transplantation, and so may be used to infer effects of transplantation on cancer outcomes. This approach has continued to be used.[14 15 16] Initially MR was defined as the use of germline genetic variation to strengthen causal inference for the influence of modifiable exposures on risk of disease or other outcomes.[17] This wider definition includes, for example, studies of gene-by-covariate interaction (often with environment as the covariate), for which the interaction cannot be viewed as an instrument for the exposure of interest.[18 19] Other study designs, such as twin studies, also use the basic principles of Mendelian genetics and so can be considered a form of MR. One such example used a male co-twin as an indicator of (on average) higher antenatal testosterone to appraise the effect of testosterone on neurodevelopmental traits.[20] MR studies range from a simple test of a SNP-outcome association, which can provide evidence as to whether an exposure affects a disease, to a specific effect estimate from an instrumental variable analysis.

The STROBE-MR guidelines are aimed at the (currently) large majority of MR studies that are implemented within an instrumental variable framework. For MR studies that do not use an instrument for the exposure (such as those of gene-by-environment interaction) or MR studies that use genetic variants in an IV framework but do not report instrumental variable estimates (such as those of sibling compatibility for transplant), some items of STROBE-MR will not be applicable, but the checklist still provides useful guidance. The table below gives an overview of study designs addressed and not addressed by STROBE-MR.

Study types addressed	Study types not addressed
One-sample MR studies	Genome-wide association studies
Two-sample MR studies	Sequencing studies
MR studies following a genome-wide association study and reported in the same article	Expression studies
One- or two- sample MR studies with multiple exposures and/or multiple outcomes	Traditional observational epidemiology studies
Partially applicable to MR studies not utilising genetic variants as instruments for an exposure and those not reporting instrumental variable approaches	

Strengthening the reporting of MR studies

Despite the growth in MR applications and methods and the increasing relevance of MR findings, there is little consensus around the reporting of MR studies. As a result, the quality of reporting of these studies has been inconsistent. Empirical evidence [21] [22] [23] indicates that many reports of MR studies do not clearly state or examine the various assumptions of MR methods and report insufficient details on the data sources, which makes it hard to evaluate the quality and reliability of the results.

The Strengthening the Reporting of Observational studies in Epidemiology (STROBE) guidelines for observational research[24] [25] were developed for the three main study designs in epidemiology (cohort, case-control and cross-sectional studies). Some of items in STROBE are either too general or not applicable to MR studies, while other items relevant to MR studies are missing. To improve reporting MR studies, we developed a separate checklist of items motivated by the STROBE guidelines but explicitly focused on the MR study design: The Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR) Statement (Table 1). Similar to the STROBE checklist, the items in STROBE-MR relate to the Title, Abstract, Introduction, Methods, Results and Discussion sections of articles. The STROBE-MR statement has been published recently.[26]

Box 2. Glossary of commonly used terms in Mendelian Randomisation

Term	Explanation
Mendelian Randomisation (MR)	A method that uses genetic variation to strengthen causal inference regarding modifiable exposures influencing risk of disease or other outcomes. The majority of MR studies are implemented within an instrumental variable (IV) framework, using genetic variants as instrumental variables.
One sample MR	A type of MR study in which a single sample of individuals is used to estimate the genetic variant-exposure and genetic variant-outcome associations. Alternatively, external weights estimating the effect of the genetic variant on the exposure may be used. This approach requires that the genetic variants, exposures and outcomes are all measured in the same sample and individual-level data are available on all participants.
Two sample MR	A type of MR study in which the genetic variant-exposure and genetic variant-outcome associations are estimated in different samples and combined using meta-analysis tools. This approach requires summary-level statistics of the association of each genetic variant in the two samples. It does not require individual-level data.
Bidirectional MR	A type of MR study in which one set of instrumental variables is used to test the effect of the exposure on the outcome and a separate set of instrumental variables is used to test the effect of the outcome on the exposure. This allows for a better understanding of the direction of the causal effect.
Instrumental variables (IV)	Variables associated with the exposure of interest that are not related to confounders, and that affect the outcome only through the exposure.
IV assumptions (core assumptions in MR studies)	*Relevance assumption*: The genetic variants are associated with the exposure of interest. *Independence assumption*: The genetic variants share no unmeasured cause with the outcome.

	Exclusion restriction assumption: The genetic variants do not affect the outcome except through their potential effect on the exposure of interest.
Assessment of IV assumptions	Various tests can assess plausibility of IV assumptions (e.g., a test of whether potential confounders or pleiotropic mechanisms are associated with the genetic variant; see Box 4 for more examples). Only the first IV assumption (*relevance*) can be tested conventionally; the validity of the other assumptions cannot be guaranteed. However, tests can provide evidence that they are unlikely to hold (i.e. these assumptions cannot be verified, but sometimes can be falsified).
Gene-environment equivalence	The notion that differences in an exposure induced by genetic variation will produce the same downstream effects on health outcomes as differences in the exposure produced by environmental influences.
Genetic variant	A variation in the DNA sequence that is found within a population. Typically, a single nucleotide polymorphism (SNP).
Single nucleotide polymorphism (SNP)	A genetic variant in which a single base pair in the DNA varies across the population, at an appreciable frequency. SNPs typically have two alleles (e.g. adenine, cytosine, guanine, or thiamine). If the SNP is associated with the trait, then one allele will be associated with a higher value of the trait, the other with a lower value. In MR studies, SNPs are the most common genetic variants used as instrumental variables for a modifiable exposure.
Strand alignment	This ensures that the alleles in the exposure GWAS and the outcome GWAS are measured on the same DNA strand. This is an issue if the SNPs are palindromic (i.e. guanine/cytosine and adenine/thymine SNPs) which would look the same on both DNA strands. Without ensuring that the exposure and outcome GWAS report the same strand, such SNPs can introduce ambiguity as to whether both the exposure and outcome GWAS are reporting the association with the same effect allele.
Allele score	A single variable produced by combining information from several SNPs that are associated with a trait or phenotype (for example blood pressure), which can be used to predict the exposure in a MR study. An allele score is sometimes also referred to as genetic risk score (GRS), polygenetic risk score (PRS), polygenic score (PGS), genetic prediction score, etc.
Linkage disequilibrium	The non-random association of alleles at two or more loci. This normally occurs within a small region of the genome in the general population. This is a potential source of bias in MR studies.
r^2	A measure of the linkage disequilibrium between two genetic loci to quantify their correlation, accounting for minor allele frequency. (A value of one denotes perfect correlation). This should not be confused with R^2 value representing the proportion of variation in the exposure variable explained by the genetic variant, which can be used to calculate instrument strength.
Test of instrument strength	Measure of association of the genetic variant and the exposure. The strength is typically tested using the partial F-statistic, the beta coefficient of the genetic variant-exposure association, or the R^2.
Test for difference	Used to assess the difference between the multivariable adjusted phenotypic association and Mendelian randomisation estimates, e.g., the Hausman test. These tests indicate whether there is any evidence that the estimates differ, over and above estimation error.
Horizontal pleiotropy	A situation in which genetic variants affect the outcome via pathways independent of the exposure. This is a violation of the *exclusion restriction* assumption and a source of bias in MR studies.

Weak instrument bias	If genetic variants used as instrumental variables are only weakly associated with the exposure of interest, they are said to be "weak instruments". A common rule is that if the partial F-statistic is less than 10 in the single sample analysis, it may exhibit weak instrument bias. If instruments are weakly associated with the exposure, then the MR estimates can be biased.
Collider bias	Bias that can occur when conditioning on a common effect of genetic variant and another key variable, such as the outcome or a confounder. This conditioning can either occur statistically (e.g. including a covariate that is caused by both the variant and outcome) or through the study sampling (e.g. analysing a sample of hospitalized patients, where hospitalisation is a consequence of the variant and outcome).
Data	Can refer to either individual-level data, e.g. measurements of participants' phenotypes such as BMI and genetic data, or SNP-level phenotype association estimates (summary-level data).

Development, scope and intended use of STROBE-MR

Described in detail elsewhere,[26] we established this initiative in 2018, following guidance for the development of medical research reporting guidelines.[27] We invited a group of experts, ranging from MR methodologists and authors of previous reporting guidelines to frequent MR study design users and scientific journal editors to participate in a workshop. The group met for a two-day face-to-face meeting in Bristol, UK, in May 2019 to discuss the empirical evidence on reporting quality of MR studies and draft the checklist items. The draft checklist was published as a preprint in July 2019,[28] and debated on the preprint platform, social media, and a dedicated session at the 4th International Mendelian Randomisation Conference.[29] The checklist was revised in the light of the comments received and an article presenting the STROBE-MR statement written.[26]

The STROBE-MR reporting guidelines are meant to apply to studies that use properties of germline genetic variants to strengthen causal inference regarding possible effects of potentially modifiable exposures on outcomes. The two principal types of MR studies are one-sample MR and two-sample MR. In a one-sample MR study, the relationships between the genetic variant and exposure and between the genetic variant and outcome are both measured in the same sample. In a two-sample MR study, these two relationships are measured in separate samples. A two-sample MR analysis can be conducted with summary-level statistics only, using associations of the genetic variants with the exposure and with the outcome obtained from the two samples. There is a continuum between one-sample and two-sample studies wherein sometimes external weights of the relationship between the genetic variants and the exposures can be used in one-sample studies. **Box 1** outlines the study designs that are covered by the STROBE-MR guidelines and study designs specifically not covered by these guidelines.

Purpose of this article

This Explanation and Elaboration (E&E) document is intended to complement the STROBE-MR statement.[26] The format follows that of previous reporting guidelines such as the

STROBE E&E document;[25] it aims to provide readers with a detailed explanation supporting each of the 20 items in the checklist and examples of transparent reporting. Examples of quality reporting for each checklist item have been identified from published MR studies.

This document should be considered as a reference to understand better what is meant by each item in the accompanying checklist. The examples do not necessarily represent the ideal statement for each checklist item. Rather, they highlight the intended issue meant to be covered in each item in the checklist. **Boxes 1-5** contain more theoretical background pertaining to MR study designs and complement recommendations on reporting. Additional guidance on performing MR studies can be found elsewhere.[30]

Some examples accompanying the items were edited by removing citations or replacing sections not related to the reported item by [...]. Items are divided into sections: Title and Abstract (item 1), Introduction (items 2-3), Methods (items 4-9), Results (items 10-13), Discussion (items 14-17) and Other Information (items 18-20). Some items have several parts (e.g., a-d) that relate to the same topic. Item 10.d only relates to a two-sample MR study design. Additional examples are provided in the Web Appendix of this document and indicated by ***WA** where available. We advise authors to address all items in the checklist, even if some information is reported in their supplementary materials due to space restriction.

Table 1. STROBE-MR checklist of recommended items to address in reports of Mendelian Randomisation studies

Item No.	Section		Checklist item
1	**TITLE and ABSTRACT**		Indicate Mendelian randomisation as the study's design in the title and/or the abstract if that is a main purpose of the study
	INTRODUCTION		
2	**Background**		Explain the scientific background and rationale for the reported study. What is the exposure? Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question
3	**Objectives**		State specific objectives clearly, including pre-specified causal hypotheses (if any)
	METHODS		
4	**Study design and data sources**		Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:
		a)	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available
		b)	Participants: Give the eligibility criteria, and the sources and methods of selection of participants
		c)	Describe measurement, quality control and selection of genetic variants
		d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases
		e)	Provide details of ethics committee approval and participant informed consent, if relevant
5	**Assumptions**		Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis
6	**Statistical methods: main analysis**		Describe statistical methods and statistics used
		a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)
		b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected
		c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples
		d)	Explain how missing data were addressed
		e)	If applicable, indicate how multiple testing was addressed
7	**Assessment of assumptions**		Describe any methods or prior knowledge used to assess the assumptions or justify their validity
8	**Sensitivity analyses and additional analyses**		Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)
9	**Software and pre-registration**		

		a)	Name statistical software and package(s), including version and settings used
		b)	State whether the study protocol and details were pre-registered (as well as when and where)
	RESULTS		
10	**Descriptive data**		
		a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow-diagram
		b)	Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions)
		c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies
		d)	For *two-sample* Mendelian randomisation: i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples ii. Provide information on the number of individuals who overlap between the exposure and outcome studies
11	**Main results**		
		a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale
		b)	Report a causal effect estimate between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per standard deviation difference
		c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
		d)	Consider plots to visualise results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)
12	**Assessment of assumptions**		
		a)	Report the assessment of the validity of the assumptions
		b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value)
13	**Sensitivity analyses and additional analyses**		
		a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions
		b)	Report results from other sensitivity analyses or additional analyses
		c)	Report any assessment of direction of causality (e.g., bidirectional MR)
		d)	When relevant, report and compare with estimates from non-MR analyses
		e)	Consider additional plots to visualise results (e.g., leave-one-out analyses)
	DISCUSSION		
14	**Key results**		Summarise key results with reference to study objectives

15	**Limitations**		Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them
16	**Interpretation**		
		a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison to other studies
		b)	Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable
		c)	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions
17	**Generalizability**		Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure
	OTHER INFORMATION		
18	**Funding**		Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based
19	**Data and data sharing**		Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where
20	**Conflicts of Interest**		All authors should declare all potential conflicts of interest

Items with Examples and Explanations

1. TITLE and ABSTRACT

Indicate Mendelian randomisation as the study's design in the title and/or the abstract if that is a main purpose of the study.

Title

Example: "BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization." [31]

Example: "Genome Wide Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders." [32]

Explanation: When MR has played a crucial role in the study design, the term "Mendelian randomisation" should be included in the title. In some situations, MR is used as a follow-on analytic technique, when the primary analysis is not MR. In this case, there may be no need to directly include MR in the title but retain focus on the manuscript's main objectives.

Abstract

Example: "**Importance** Human genetic studies have indicated that plasma lipoprotein(a) (Lp[a]) is causally associated with the risk of coronary heart disease (CHD), but randomized

trials of several therapies that reduce Lp(a) levels by 25% to 35% have not provided any evidence that lowering Lp(a) level reduces CHD risk.

Objective: To estimate the magnitude of the change in plasma Lp(a) levels needed to have the same evidence of an association with CHD risk as a 38.67-mg/dL (ie, 1-mmol/L) change in low-density lipoprotein cholesterol (LDL-C) level, a change that has been shown to produce a clinically meaningful reduction in the risk of CHD.

Design, Setting, and Participants: A Mendelian randomization analysis was conducted using individual participant data from 5 studies and with external validation using summarized data from 48 studies. Population-based prospective cohort and case-control studies featured 20 793 individuals with CHD and 27 540 controls with individual participant data, whereas summarized data included 62 240 patients with CHD and 127 299 controls. Data were analyzed from November 2016 to March 2018.

Exposures: Genetic *LPA* score and plasma Lp(a) mass concentration.

Main Outcomes and Measures: Coronary heart disease.

Results: Of the included study participants, 53% were men, all were of white European ancestry, and the mean age was 57.5 years. The association of genetically predicted Lp(a) with CHD risk was linearly proportional to the absolute change in Lp(a) concentration. A 10-mg/dL lower genetically predicted Lp(a) concentration was associated with a 5.8% lower CHD risk (odds ratio [OR], 0.942; 95% CI, 0.933-0.951; $P = 3 \times 10^{-37}$), whereas a 10-mg/dL lower genetically predicted LDL-C level estimated using an LDL-C genetic score was associated with a 14.5% lower CHD risk (OR, 0.855; 95% CI, 0.818-0.893; $P = 2 \times 10^{-12}$). Thus, a 101.5-mg/dL change (95% CI, 71.0-137.0) in Lp(a) concentration had the same association with CHD risk as a 38.67-mg/dL change in LDL-C level. The association of genetically predicted Lp(a) concentration with CHD risk appeared to be independent of changes in LDL-C level owing to genetic variants that mimic the relationship of statins, PCSK9 inhibitors, and ezetimibe with CHD risk.

Conclusions and Relevance The clinical benefit of lowering Lp(a) is likely to be proportional to the absolute reduction in Lp(a) concentration. Large absolute reductions in Lp(a) of approximately 100 mg/dL may be required to produce a clinically meaningful reduction in the risk of CHD similar in magnitude to what can be achieved by lowering LDL-C level by 38.67 mg/dL (ie, 1 mmol/L)." [33] ***WA***

Explanation: The abstract should provide an informative and balanced summary of what was done and what was found. This should be presented alongside critical issues in study design, including sources of data, exposures/outcomes, individual vs summary data and would (if possible) include the term, "Mendelian randomisation", to make the article discoverable as such. Results should be presented in a fully transparent manner and include both point estimates and their error (i.e. not only p-values) for the range of approaches applied. The word "causal" should be used carefully, as MR only provides estimates intended to inform our understanding of causal relationships under strong assumptions. The abstract should be sufficiently detailed to act as a stand-alone part of the manuscript. When permitted by the journal, structured abstracts can provide clarity and help assure that all relevant information is included. Additional examples of abstracts for one-sample, two-sample and embedded MR studies can be found in the Web Appendix.

INTRODUCTION

2. Background

Explain the scientific background and rationale for the reported study. What is the exposure? Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.

Example: "Epidemiologic studies have reported an increased risk of multiple sclerosis (MS) with earlier age at puberty, particularly among women. However, others failed to replicate this finding. Pubertal timing has complex interactions with weight status, whereby higher childhood adiposity leads to earlier puberty, which in turn is associated with higher adult body mass index (BMI). Because evidence supports a role for increased BMI in MS pathogenesis, at least part of the observed link between pubertal timing and MS might be explained by BMI. Some of the limitations faced by observational studies can be mitigated through instrumental variable methods, in which a variable is used as a proxy for an exposure to explore the effect of that exposure on an outcome. In Mendelian randomisation (MR), genetic variants are used as instrumental variables to test for a causal association between a risk factor and an outcome." [34]

Explanation: While some authors have used MR to test the effect of exposures on many different outcomes without prior hypotheses,[35] most MR studies are designed to assess a specific hypothesis that has arisen from prior studies. When the latter is the case, the rationale for assessing the current hypothesis should be described, including the *a priori* expectation of the effect size.

MR can potentially be used to test causal null hypotheses or estimate point, period, or lifetime effects. The specific role of MR in assessing the study hypothesis should be delineated to orient the reader as to what specific gap in the literature can be addressed by applying MR methods to the study hypothesis.

3. Objectives

State specific objectives clearly, including pre-specified causal hypotheses (if any).

Example: "Objective: To evaluate the potential causal association between genetic variants related to elevated serum calcium levels and risk of coronary artery disease (CAD) and myocardial infarction using Mendelian randomization." [36]

Explanation: Authors should clearly state that the study aims to estimate a causal effect of the specified exposure on the specified outcome. This section should define the key exposure(s) and outcome(s) of interest to orient the reader, and state the overall study objectives.

METHODS

4. Study design and data sources

Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study.

Example: "To assess the effect of age at puberty on MS susceptibility, we performed a 2-sample MR analysis using genetic associations from large genome-wide association studies (GWASs). Here, we first explore whether participants with genetic variants associated with later age at puberty also have a lower risk of MS. Second, we use recent developments in MR methodology to determine whether pubertal timing exerts direct effects on MS susceptibility independently of BMI. [...] To select our genetic instruments, we used single-nucleotide polymorphisms (SNPs) associated with age at menarche from the largest GWAS meta-analysis to date by the Reproductive Genetics (ReproGen) consortium, combining 329,245 women of European ancestry (table)." [34]

"Table: Details of datasets included in the MR analyses"

Table Details of datasets included in the MR analyses

Phenotype	Consortium	Participants (cases with MS), n	Source (if publicly available)
MS	IMSGC	41,505 (14,802)	Not available
Age at menarche	ReproGen	329,345	reprogen.org/data_download.html
Adult BMI (age, sex adjusted)	ENGAGE	87,048	diagram-consortium.org/2015_ENGAGE_1KG/
Childhood BMI (age, sex adjusted)	EGG	35,668	egg-consortium.org/childhood-bmi.html

Abbreviations: BMI = body mass index; EGG = Early Growth Genetics; ENGAGE = European Network for Genetic and Genomic Epidemiology; IMSGC = International Multiple Sclerosis Genetics Consortium; MR = mendelian randomization; MS = multiple sclerosis; ReproGen = Reproduction Genetics.

Table reproduced with permission from Harroud et al., 2019.[34]

Explanation: As in STROBE,[25] presenting critical elements of study design early in the article allows readers to orient themselves on the study basics. Authors should clarify whether the MR study used individual-level participant data or SNP-level summary data, i.e. whether it uses a one-sample or two-sample MR design. Some MR studies draw on multiple sources of data (e.g. different sources for the ascertainment of the association between the genetic variant and exposure, and for the association between the genetic variant and outcome). Furthermore, sources of data may be from meta-analyses of multiple samples. The general design and data sources should therefore be made clear.

We recommend a table to provide clear documentation of the sources of genetic-variant level information for the MR study (see example Table). The genetic variants used to estimate the exposure may have been ascertained in one study, but the effect size (or weight) of these genetic variants on the exposure taken from a separate study. If this is the case, we recommend reporting both sources of information. The table should be expanded as required. For example, if different MR studies with different outcomes are added to the study, then authors should add additional columns to the table. If additional exposures are studied, then additional rows can be added.

If data were extracted from pre-existing studies, describe how the data were obtained. If data are publicly available, provide a hyperlink to the data source, where possible. If using summary-level data, ensure all of these details are traceable and allow for a qualitative assessment of data sources' heterogeneity.

For each data source contributing to the analysis, describe the following:

4a) Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.

Example: "This study comprised a meta-analysis of directly genotyped and imputed SNPs from 21 cohorts totalling 42,024 individuals (Table 1). An expanded description of the participating studies is provided in the Text S2." [37]

Example: "A total of 23 cohorts with genome wide genotyping and fracture data were recruited globally through the GEnetic Factors for OSteoporosis consortium (GEFOS; http://www.gefos.org/). These cohorts were predominantly of European descent and from Europe (n=13), North America (n=8), Australia (n=1), and east Asia (n=1; tables S1A and S2A), and included 20 439 fracture cases and 78 843 controls." [38]

Explanation: Readers need information on the population(s) studied, setting, and locations to assess the context and generalisability of the study results. Exposures such as environmental factors and therapies can change over time. Also, study methods may evolve over time. Knowing when a study took place and over what period participants were recruited and followed up puts the study in historical context, which is essential for interpreting results. Where such information has been described in previous publications, unambiguous reference to these is likely to be sufficient. Providing a description of the ancestry of the participants will help understand potential sources of heterogeneity and generalizability of results. If using summary-level data from existing studies, ensure details are traceable to allow for a qualitative assessment of any heterogeneity of settings across data sources.

4b) Participants: Give the eligibility criteria, and the sources and methods of selection of participants.

Example: "The UK Biobank recruited more than 500 000 people aged 37-73 years (99.5% were between 40 and 69 years) from across the country in 2006-10. Participants provided a range of information via questionnaires and interviews (such as demographics, health status, and lifestyle); anthropometric measurements, blood pressure readings, and blood, urine and saliva samples were taken for future analysis. This has been described in more detail elsewhere. We used 120 286 participants of white British descent from the initial UK Biobank dataset, of whom 119 669 had valid genetic data and both BMI and height measures available. We did not include other ethnic groups, because individually they were underpowered." [39]

Explanation: Detailed descriptions of the study participants and sampling frame help readers understand the applicability of the results. Authors should provide all the eligibility criteria, sources and methods for selecting participants, and methods of follow up where applicable. Method of recruitment into the study should likewise be described. Where such information has been described in previous publications, unambiguous reference to these is likely to be sufficient. If using summary-level data from existing studies, ensure details are traceable to allow for a qualitative assessment of participants' heterogeneity across data sources.

In case-control studies, the choice of cases and controls is crucial to interpreting the results, and the method of their selection has major implications for study validity. In general, controls should reflect the population from which the cases arose.

4c) Describe measurement, quality control and selection of genetic variants.

Example: "Genotyping was conducted using the Affymetrix UK Biobank Array. Autosomal analysis was restricted to up to 13,977,204 high-quality Haplotype Reference Consortium imputed variants with a MAF > 0.05%, minor allele count > 5, info score > 0.3, genotype hard call rate > 0.95, and Hardy–Weinberg $P > 1 \times 10^{-6}$." [40]

Example: "Genetic markers for various obesity-related risk factors comprised SNPs that were associated with the risk factor of interest ($P < 5 \times 10^{-8}$) based on study participants with European ancestry. Correlated SNPs were excluded based on measures of linkage disequilibrium (LD) R2 < 0.1. [...] SNPs with ambiguous strand codification (A/T or C/G) were replaced by SNPs in genetic linkage (R2 > 0.8) using the proxy snps R package (European populations) (R Project) or were removed from the analyses if the minor allele frequency was higher than 0.4." [41]

Explanation: Providing information on the ascertainment of genotypes and their quality control will enable readers to assess the quality of the genetic variants used in the study. In the case of two-sample MR, this will often require referring to supplementary material presented in previously published articles.

The methods section should provide a clear explanation of the selection and inclusion of specific genetic variants in the analysis. This would include a description of genetic variants allocated to the exposure of interest and in the case of reverse MR, the genetic variants allocated to the outcome. For each variant, the "rsID" or base position and chromosome should be disclosed along with clear reasoning for the variant choice and, additionally, the reference panel used. The reasoning could include the evidence of association with the exposure or outcome of interest *or* the characteristics that qualify the specific variant to be used, such as linkage disequilibrium (LD) in the case of proxy use. The inclusion of variants in high LD may not contribute additional information in estimating the causal effect. It may even lead to biased estimates of standard errors if this correlation structure is not accounted for.[42] Authors should define the threshold used to select independent variants (e.g. the r^2), the reference panel and the population under investigation. However, there are cases in which variants (although in LD) from a specific gene region with biological relevance regarding the exposure of interest can also be included. In this situation, the authors need to describe in which biological pathways these variants are implicated, the r^2 threshold for inclusion, and which method was used to model the correlation structure.

Further information is required to chart the management of genetic variants and harmonisation of data sets in two-sample MR analysis. This includes the conditions used to identify proxy variants in the absence of the same variant being available in both datasets (e.g. LD threshold), the presence/absence and handling of strand alignment, and orientation of effect and non-effect alleles. Other aspects, such as temporal stability of association, population specificity, or biological priors, may help understand the selected genetic variants' validity to be used as instrumental variables.

4d) For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases.

Example *(continuous exposure or outcome):* "Outcomes of the study were WHR (stages 1 and 2b), hip and waist circumference (stage 2a), compartmental body fat masses (stage 3), [...] WHR was defined as the ratio of the circumference of the waist to that of the hip, both

of which were estimated in centimeters using a Seca 200-cm tape measure. [...] Compartmental fat masses were measured in grams by DEXA, a whole-body, low-intensity x-ray scan that precisely quantifies fat mass in different body regions [...] using a Lunar Prodigy advanced fan beam scanner (GE Healthcare). Participants were scanned by trained operators using standard imaging and positioning protocols. All images were manually processed by one trained researcher, who corrected DEXA demarcations according to a standardised procedure […]." [43] ***WA***

Explanation: This section provides details on the choice and definition of key exposures, outcomes, and confounders used in the analyses. Where several outcomes or hypothesis-free approaches are used, this should be clearly indicated, together with any method that accounts for multiple testing. This section will ideally include definitions used in each study (for meta-analyses of different studies) or provide a brief summary with a clear reference if this has previously been described for the study sample. This allows readers to consider the case definition's sensitivity and specificity, and assess the relevance for their question or generalisability to their population of interest.

4e) Provide details of ethics committee approval and participant informed consent, if relevant.

Example: "Informed consent was obtained from all participants, and study protocols were approved by the local, regional, or institutional ethics committees." [44]

Explanation: All investigators must ensure that the planning, conduct, and reporting of human research are in accordance with the Helsinki Declaration as revised in 2013.[45] Authors need to provide information on the approval from the responsible ethics committee and acquisition of the informed consent. This information should also be made available if the data were obtained from publicly available sources or previously conducted studies. Authors should make sure that their study falls into the scope of the original ethics committee approval and does not violate the original agreement.

5. Assumptions

Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well as assumptions for any additional or sensitivity analysis.

Example: "As in any Mendelian randomisation analysis, several assumptions were made, including that the genetic instruments were associated with the risk factor of interest, were independent of potential confounders, and could only affect the outcome through the risk factor and not through alternative pathways (that is, through pleiotropy)." [46]

Example: "Additionally, the slope of MR-Egger regression can provide pleiotropy-corrected causal estimates; [...] An important condition of approach is that a SNP's association with the exposure variable must be independent of its direct effects upon the outcome (previously described as the InSIDE assumption)." [47]

Explanation: Explicitly stating the three core IV assumptions (**Box 3**), ideally in the methods sections, can help readers understand the underlying premises of the MR method, and to allow them to judge their validity. Ideally, the assumptions would be stated in the text using

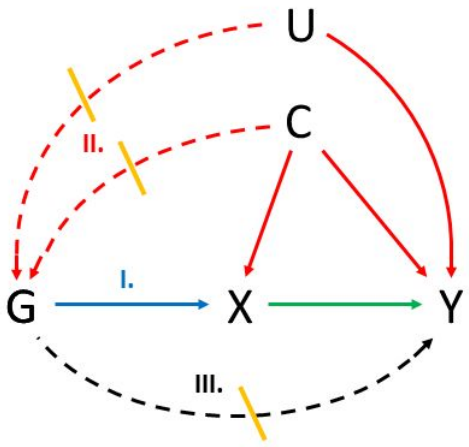
intuitive language specific to the study setting and what they imply in the context of the question being asked. Articulating the assumptions also motivates sensitivity analyses and other additional analyses used to assess the assumptions or the robustness of conclusions to their violations.

When instrumental variable (IV) estimation is used to obtain estimates of the causal effect, then a fourth assumption should be stated: typically, an assumption of effect homogeneity[48] or monotonicity.[49] In many MR studies, other methods are used to augment traditional instrumental variables estimation, and their assumptions should also be stated. For example, MR Egger regression[50] or weighted median[51 52] are often used as a supplementary analysis to obtain the estimates when multiple genetic variants are used. For more details on IV assumptions, common violations, and assessment, see **Box 3** and **Box 4**.

Box 3. Instrumental variables (IV) assumptions and Mendelian randomisation

Core IV estimation assumptions and additional assumptions	Most MR studies rely on three core IV assumptions (*relevance*, *independence* and *exclusion restriction*), as described in Box 2, to carry out testing for causal effects of the exposure on the outcome.[17] Estimating effect sizes through the instrumental variable approach imposes a fourth assumption, usually homogeneity of effects of the exposure on the outcome.[17 53 54 55] The homogeneity assumption can also be replaced by imposing a monotonicity assumption that an increase in the number of risk alleles never, for any individual, lowers the likelihood of exposure, typically leading to estimating an effect in a subgroup of the study population.[56]
Violations	The exclusion restriction is sometimes also referred to as an assumption of "no horizontal pleiotropy" (see Box 2), but can be violated in several other ways (e.g. by gene-exposure interaction, by having some forms of time-varying exposures, or by measurement error in the exposure or a multi-component exposure).[54] Concerns about violations of the independence assumption usually focus on confounding by ancestry (or population stratification). However, it can also be violated by various forms of selection/collider bias, by dynastic effects, or by assortative mating.[57 58 59] When multiple variants are used in the analysis, these assumptions pertain to each of the variants. Other methods can relax these assumptions, as described below.
Assumptions for additional analyses	In many MR studies, traditional instrumental variables estimation methods have been extended in several ways. For example, when multiple genetic variants are used, MR Egger regression[50] or weighted median[51 52] are often used as a supplementary analysis to obtain estimates. MR Egger regression relaxes the exclusion restriction assumption but itself imposes an "InSIDE" assumption that "the size of the direct effects of the genetic variants on the outcome that do not operate through the exposure are independent of the size of the genetic variants' effects on the exposure." Additionally, the two-sample MR approach assumes that the association between the genetic variants and the exposure is the same in the two samples, which may not hold if samples are selected from different subpopulations (e.g. by sex, age, ethnicity).

Box 3 Figure: A Directed Acyclic Graph (DAG)# Illustrating the Assumptions of Instrumental Variable Analyses. A genetic variant G is used as an instrumental variable (proxy) for the exposure X to assess its causal effect on the outcome Y.



Instrumental Variable (IV) Assumptions:
I. Relevance: The genetic variant (G) is associated with the exposure of interest (X)
II. Independence: The genetic variant (G) shares no unmeasured cause with the outcome (Y)
III. Exclusion restriction: The genetic variant (G) does not affect the outcome (Y) except through its potential effect on the exposure of interest (X)

Solid arrows indicate causal effects; e.g. X → Y indicates a causal effect of the exposure X on the outcome Y. Dashed arrows indicate causal effects that are specifically prohibited by the IV assumptions.

Box 4. Assessment and falsification of the assumptions for Mendelian Randomisation

Relevance	For the *relevance* assumption, authors should report how they measured instrument strength. Reporting the F-statistic, if individual-level data are available, provides several advantages for understanding the risk of weak instrument bias.[60] The F-statistic can also be approximated using summary-level data. If the proposed instrument strength is low, reporting should include whether approaches that are robust to weak instruments have been used.
Independence	The *independence* assumption cannot be directly verified, but it can be partially assessed in many research settings. Negative control outcomes or negative control populations can sometimes evaluate the reasonableness of the assumption.[61] Reporting associations between measured covariates that may confound the variant-outcome relationship can also prove helpful, particularly if scaled by instrument strength[62 63] or presented alongside an E-value or related bias analytic approach.[64]
Exclusion restriction	For the *exclusion restriction* assumption, MR Egger regression[50] can be used to detect certain versions of pleiotropy and therefore provide evidence of certain violations of the exclusion restriction. However, the approach depends on an additional assumption (described above) and requires multiple independent variants. Additional approaches to test the exclusion restriction include weighted median[51] and mode.[52] The use of negative control outcomes or negative control populations can also sometimes allow evaluation of this assumption.[65] The use of known biologic effect of a SNP can also be leveraged to decrease the probability of violation of this assumption.
Homogeneity	The homogeneity assumption, if used, is also not directly verifiable. One possibility for supporting its validity is to address whether the effect estimate, or even the genetic variants' effects on the exposure, is the same across subpopulations.[62 66] Authors can perform stratified or adjusting analyses to relax this assumption if meaningfully different effects are estimated in different subpopulations.[67] Furthermore, a global exploration of the homogeneity assumption can examine if there are differences in variance of a continuous outcome across the genetic instrument; the extent of such differences provides evidence as to extent of violation of the homogeneity assumption.
Joint falsification strategies	Some falsification strategies assess assumptions jointly. When using multiple genetic variants as proposed instruments, it is possible to test whether heterogeneity exists across the separate effect estimates (see "Test for difference" in Box 2). Although this test is often interpreted as an assessment of the exclusion restriction, it is jointly testing the exclusion restriction, independence, and homogeneity assumptions. Another relatively straightforward joint test of all assumptions comes from comparing the effect estimate with that obtained using a more traditional confounding-adjustment approach.[68] Assuming the traditional approach is biased due to unmeasured confounding, and the direction of that confounding is suspected, examining whether the MR effect estimate aligns with the suspected direction of confounding can support the joint validity of the assumptions underlying the MR effect estimate.
Sensitivity analyses	Because several of the estimators using multiple genetic variants rely on different versions of relaxing or adapting the instrumental variable assumptions (e.g., MR Egger, median-based or modal-based estimators), a comparison of estimates obtained using each of these approaches can help understand the sensitivity of effect estimates to the non-overlapping assumptions of each.[69] Researchers may also compare MR effect estimates to non-MR estimates, depending on the assumptions underlying alternative methods. Independent replication of MR findings in an independent dataset or with a different study design (e.g. one-sample vs two-sample MR) is typically advocated to assess the findings' robustness. Many of the traditional bias analytic techniques in epidemiology can be adapted for MR readily, including the E-value[64 70] or formulas for understanding the magnitude and direction of confounding bias[62 63] or for violations of the exclusion restriction.[54 71] When selection bias is a concern, researchers also frequently conduct simulations to understand the plausible size and direction of bias.[59] Simulations may also help understand the plausible size and direction of bias induced by assortative mating,[57] dynastic effects,[9 58] and time-varying effects,[72] if deemed relevant.

6. Statistical methods: main analysis

Describe statistical methods and statistics used.

6a) Describe how quantitative variables were handled in the analyses (i.e., scale, units, model).

Example: "The effect size for each meta-analysis is reported in the main results as the effect of a one-standard-deviation (1-SD) change in natural-log-transformed 25OHD level, since this metric is more interpretable than an arbitrary difference…. In order to provide a better clinical interpretation of a 1-SD change in natural-log-transformed 25OHD level, we selected three different clinically relevant 25OHD thresholds for vitamin D status (<25 nmol/l for vitamin D deficiency, <50 nmol/l for vitamin D insufficiency, and >75 nmol/l for vitamin D sufficiency)." [73]

Explanation: Any transformations made in the quantitative variables (i.e. exposure, outcome, or the relevant covariates) should be explicitly mentioned, as this affects both interpretation of results and their comparability with other studies. Describing biological knowledge or prior evidence can help justify chosen groupings. When possible, authors should also back-transform estimates in efforts to report the units of measurement in common terms to enable future replication of findings.

6b) Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected.

Example: "We created an allele score from 97 genetic variants previously found to be associated with BMI, in a recent GWAS meta-analysis by the GIANT consortium. 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 such that a higher allele score corresponds to a higher BMI, and was standardised to have a mean of zero and SD of 1)." [74]

Explanation: An allele score (also sometimes referred to as genetic risk score, polygenic risk scores, genetic prediction scores, etc.) is a variable that summarises multiple genetic variants in a single measure. When many variants are included in the score, bias and coverage probabilities of the IV estimates are improved compared with estimates from the two-stage least squares approach.[75] It is considered good practice to explicitly define the criteria for selecting variants included in the allele score and whether these are based on external data. An allele score could be weighted or unweighted. If weighted, it should be clarified if the weights are derived from the data under analysis or an independent data source. Authors should also report which genetic model of inheritance is implied in the calculation of genetic variant-exposure and variant-outcome associations (i.e., additive or multiplicative).

6c) Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples.

Example: "Genetic associations with all exposures were taken from a large meta-analysis of GWAS, conducted in adults (n = 108,557; mean age, 50.6 years; ~53% men) of European ancestry, without diabetes, adjusted for age, sex, study site and geographic covariates using an additive genetic model. [...] Genetic associations with MI, angina and heart failure were

obtained using logistic regression controlling for age, assay array and 10 principal components in sex-specific analysis and additionally adjusted for sex in the overall analysis, as the adjustment in our previous MR study in the UK Biobank. [...] Specifically, we obtained SNP-specific Wald estimates (quotient of genetic association on outcome and genetic association on insulin) and then meta-analysed them using inverse variance weighting (IVW) with multiplicative random effects." [76]

Explanation: The authors should present all the analytical details on the calculation of the IV estimator. Further clarification on estimating the associated standard errors should also be provided (i.e. if this is based on a normal approximation, bootstrapping or other approaches). Covariates used in the MR analysis should be detailed. For a two-sample analysis, covariates used in the estimation of the genetic variant-exposure and genetic variant-outcome associations should be presented to assess whether any imbalance in covariates' use may lead to a bias.

6d) Explain how missing data were addressed.

Example: "We conduct our analyses in a Bayesian framework as this lends itself naturally to data imputation. We first introduce a Bayesian complete-case analysis method and then 4 methods for imputing data under the missing-at-random assumption that can be incorporated into the Bayesian model to include subjects with missing data […]. We use cross-sectional baseline data on 3,693 participants who have complete or partial data for C-reactive protein, fibrinogen, and the 3 SNPs. There is missingness in 2.1% of participants for C-reactive protein, 2.4% for fibrinogen, 10.8% for rs1205, 1.9% for rs1130864, and 2.6% for rs1800947." [77]

Explanation: The inclusion of multiple variants with missing data in the estimation of causal effects may decrease precision. The percentage of missing data should be presented as well as whether any imputation was performed.

6e) If applicable, indicate how multiple testing was addressed.

Example: "The significance threshold for all cancer risk and mortality is 0.004 (6 PUFAs times 2 outcomes (risk, mortality) require correcting 0.05 by 12 tests). […] Given 6 individual cancers were considered, we set a significance threshold of 0.004/36 = 0.0001." [78]

Explanation: In an MR analysis that involves multiple exposures or multiple outcomes, the authors should state whether and how they accounted for multiple testing and provide justification. They should state whether the correction was for the total number of statistically independent exposures/outcomes for all exposures and outcomes. Such a correction could involve reporting false discovery rates, Bonferroni correction, or other techniques, as outlined in the above example.

7. Assessment of assumptions

Describe any methods or prior knowledge used to assess the assumptions or justify their validity.

Example: "Mendelian randomisation was implemented using the two stage least squares method in the R package *ivpack*. We included age and sex as covariates. To assess the risk

of weak instrument bias, we used F tests to determine the strength of association in the first stage regressions between allele score and exposure. […] We used confounding bias plots to assess relative bias in the instrumental variable estimate compared with standard multivariable regression. […] To investigate the degree of bias in the initial causal estimates due to pleiotropic effects, we used two sensitivity analyses (mendelian randomisation-Egger and weighted median mendelian randomisation). [...] Mendelian randomisation-Egger and weighted median methods were implemented using the R package TwoSampleMR." [79]

Example: "There are reasons for considering MR analysis of a protein drug target to be a distinct category of MR analysis. […] Aside from mRNA expression, differences in protein expression or function are the most proximal consequence of natural genetic variation. This has two consequences: frequently, variants located in and around the encoding gene can be identified with a very substantial effect on protein expression in comparison to other traits; moreover such instruments may also be less prone to violating the 'no horizontal pleiotropy' assumption than variants located elsewhere in the genome […] In the case of MR analysis of proteins, Crick's 'Central Dogma' imposes an order on the direction of information flow from gene to mRNA to encoded protein, which does not extend beyond this to other biological traits that lie more distally in the causal chain that connects genetic variation to disease risk. Finally, *cis*-MR of a protein risk factor greatly reduces the risk of reverse causation, because Crick's dogma indicates that the pathway gene → encoded protein → disease would always be favoured over the pathway gene → disease → encoded protein, especially given that the gene → encoded protein association is typically derived from population-based (disease-free) samples. Thus, from an MR perspective, proteins are in a privileged position compared with other categories of risk factor and the use of *cis*-MR represents an optimal approach to instrument their causal effect for disease." [80]

Explanation: For each of the assumptions underlying an MR analysis, authors should report any methods used to assess the assumptions or justify their validity. Generally, the subject-related background can be used to support the reasonableness of each assumption. While many assumptions cannot be verified, there are methods available to attempt to falsify them. In line with the relevance assumption, the authors can report how they assessed instrument strength. If the proposed instrument strength is low, reporting might include whether approaches that are robust to weak instruments have been used. Although many possible methods are available, some are usable only in certain settings (e.g., for dichotomous exposures). **Box 4** describes some of the more common and useful approaches, and **Table 2** lists the commonly used statistics used when examining assumptions and performing sensitivity analyses. The first three core assumptions pertain to any MR analysis with a single instrument; the additional assumptions are needed for instrumental variable estimation. Exclusion restriction is relaxed in sensitivity analyses such as MR Egger regression. These assessments and sensitivity analyses do not represent an exhaustive list of possible strategies, and not all sensitivity analyses are relevant to all MR analyses. For example, F statistics are principally relevant for instrumental variable analyses, as they are approximately equivalent to variance explained in two-sample MR approaches based on GWAS output. Associations with measured covariates (e.g., age, sex and race/ethnicity and effect estimates across sub-populations can be reported with one-sample, but **generally** not with two-sample MR studies, however, GWAS summary statistics are increasingly available for sex- and ancestry-specific analyses. More comprehensive reviews are provided in Glymour et al.[68] and Labrecque and Swanson.[81]

Table 2: Summary of the most common IV assumptions and examples of their possible assessments or sensitivity analyses.

Assumptions	Examples of Possible Assessments
1. "*Relevance":* The genetic variants are associated with the exposure of interest.	Report F statistic.
2. "*Independence":* The genetic variants share no unmeasured cause with the outcome.	Report on associations of plausible confounders with both the genetic variant(s) and the outcome. Report on how population stratification has been taken into account (e.g. through principal component adjustment) Report E-values[63] for the variant-outcome association.
3. "*Exclusion restriction"*: The genetic variants do not affect the outcome except through their potential effect on the exposure of interest.	Report results from MR Egger Regression slope estimate, as well as, the intercept and their 95% confidence intervals. Report results using negative control outcomes or negative control populations.
"Homogeneity" (2-Stage Least Squares): There is a constant causal effect of the exposure of interest on the outcome.	Report on the IV effect estimate for different measurable subpopulations*. For continuous outcomes report on variance by level of instrument.[67]
"InSIDE" (MR Egger): The genetic variants' association with the exposure variable must be independent of its direct effects upon the outcome.	Report effect estimates also from other estimators that do not require this assumption (e.g., median- and modal-based) tests.

* For example, stratified by age, race/ethnicity, gender, or socioeconomic status

8. Sensitivity analyses and additional analyses

Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations).

Example: "*Confounding*: We used confounding bias plots to assess relative bias in the instrumental variable estimate compared with standard multivariable regression. Such analyses are designed to quantify the bias present in a mendelian randomisation analysis in a manner analogous to examining the effect of adjusting or not adjusting for a potential confounder in a standard regression analysis. Additionally, in supplementary analyses we included suspected confounding factors as covariates (see supplementary table 4). The confounding variables considered were the first 10 genetic principal components, Townsend deprivation index, birth weight, breast fed, and place of birth (northing and easting coordinates).

Horizontal (genetic) pleiotropy: To investigate the degree of bias in the initial causal estimates due to pleiotropic effects, we used two sensitivity analyses (mendelian randomisation-Egger and weighted median mendelian randomisation). Mendelian randomisation-Egger is not valid for studies in which the instrumental variable-exposure and instrumental variable-outcome associations are calculated in the same sample (as was done for the main analyses in this study). Therefore, we ran the mendelian randomisation-Egger as a split sample analysis, by randomly splitting the sample in half (groups A and B). The supplementary data table shows the associations of the variants and time spent in education and refractive error for each group. Mendelian randomisation-Egger and weighted median methods were implemented using the R package TwoSampleMR (github.com/MRCIEU/TwoSampleMR).

Measurement error: To ensure the association between time spent in education and myopia was not an artefact of the non-normal distribution of the variable for age when full time education was completed, we used two alternative methods to recode time spent in education: dichotomisation into age more than 16 years when education was completed and age 16 years or less when education was completed; and excluding those who attended college or university. We compared the results with the original analyses using the continuous variable for age when full time education was completed." [82]

Example: "Tests of association for individual genetic variants were complemented with gene-based tests of association and S-PrediXcan analysis. The latter was used to identify genes with differential expression levels in cannabis users versus nonusers. We further estimated the genetic correlation of lifetime cannabis use with other traits, including use of other substances and mental health traits, such as schizophrenia. Lastly, we performed bidirectional two-sample Mendelian randomization analysis to examine whether there was evidence for a causal relationship from cannabis use to schizophrenia risk, and from liability to schizophrenia to cannabis use." [83] ***WA***

Explanation: Sensitivity analyses can test the robustness of effect estimates to plausible violations of the underlying assumptions and help understand the plausible size or direction of bias. It is hence of interest to report on any such sensitivity analyses performed. Some common strategies are described in **Box 4**, and further information is available elsewhere.[50 69 81]

9. Software and pre-registration

9a) Name statistical software and package(s), including version and setting used

Example: "We performed the analysis by using Stata version 14 (StataCorp LP) and R version 3.4.3 (The R Foundation for Statistical Computing). We used the mrrobust package for Stata and the TwoSample MR package for R to facilitate MR analyses." [84]

Explanation: Statistical methods and software should ideally be described with enough detail to enable a knowledgeable user with access to the original data to verify the reported results. It is preferable to provide the statistical code used in an online repository.

9b) State whether the study protocol and details were pre-registered (as well as when and where).

Explanation: Authors should report if a study was pre-registered, and provide a link to the study protocol. Examples of pre-registration in Mendelian randomisation are rare at present, in part because it poses challenges for secondary data analysis. Solutions overcoming this that protect against researcher bias have been proposed.[85] Wider adoption of these methods should increase the accuracy, transparency and robustness of MR studies.

RESULTS

10. Descriptive data

10a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider the use of a flow diagram.

Example: "UK Biobank recruited 502 664 participants aged 40 to 69 years through 22 assessment centres across the UK. [...] All participants completed sociodemographic questionnaires, which included questions on past educational and professional qualifications. In the latter stages of recruitment, an ophthalmic assessment was introduced, and this was completed by approximately 23% of participants. [...] In total, 69 798 participants had valid education, refractive error, and genetic data available (fig 1)." [82]

"Fig 1. Numbers of participants in UK Biobank who passed validation for Mendelian randomisation study. MSE=mean spherical equivalent."

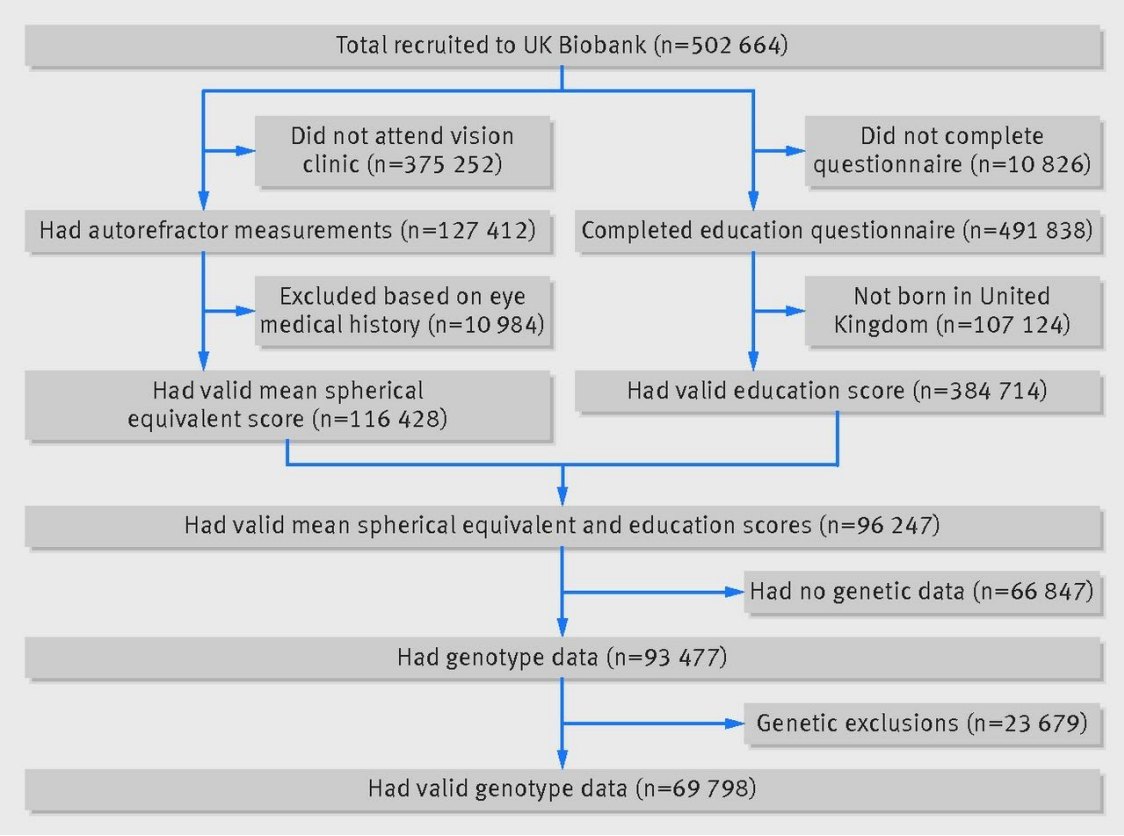


Figure reproduced with permission from Mountjoy et al., 2018.[82]

Explanation: Information on study participants will help readers understand the target population and assess the validity and generalizability of results. It also provides readers with the information needed to replicate the study and to assess whether the study is likely to exhibit collider bias. If the data sources include individual-level data, authors should report information on the participants in the study. Specifically, report the number of individuals at each stage of the study and the reasons why individuals were excluded from further study. Examples of such reasons include loss to follow-up, removal for lack of data, and quality control. Including a STROBE[25] flow chart for inclusion into the study can quickly provide information about how the study sample was selected. Where possible, report missing values for variables.

10b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions).

Example: "The UK Biobank sample comprised 53.7% women (Table 1), and the median age at recruitment was 58.0 years (interquartile range 51.0-63.0). The distribution of adiposity [exposure] and smoking behaviour [outcome] variables in the UK Biobank sample are described in table 1 and table 2. As observed in previous studies, current smokers had a lower body mass index than never smokers (−0.22 (95% confidence interval −0.27 to −0.16)). Conversely, former smokers had a higher body mass index than current smokers (1.04 (0.98 to 1.09))." [46]

"Table 1: Sample characteristics of body size parameters by smoking and sex categories in UK Biobank. Data are mean (standard deviation)."

		Smoking category			Sex	
Body size parameters	Total (n=372 791)	Never (n=203 735)	Former (n=131 537)	Current (n=37 519)	Female (n=200 247)	Male (n=172 544)
Body mass index	27.4 (4.8)	27.1 (4.7)	28.0 (4.7)	27.0 (4.8)	27.0 (5.1)	27.9 (4.2)
Weight (kg)	78.3 (15.9)	77.0 (15.6)	80.5 (16.0)	78.0 (16.3)	71.5 (13.9)	86.2 (14.3)
Height (cm)	168.8 (9.2)	168.3 (9.3)	169.4 (9.1)	169.5 (9.2)	162.7 (6.2)	175.9 (6.7)
Waist circumference (cm)	90.4 (13.5)	88.8 (13.2)	92.6 (13.6)	91.2 (13.5)	84.6 (12.5)	97.1 (11.3)
Body fat percentage (%)	31.4 (8.5)	31.5 (8.6)	31.7 (8.2)	29.9 (8.6)	36.6 (6.9)	25.3 (5.8)

"Table 2: Sample characteristics of smoking parameters by body mass index and sex categories in UK Biobank ever smokers (current plus former smokers). Data are mean (standard deviation)."

		Body mass index category				Sex	
Smoking parameters	Total (n=169 056)	Underweight (<18.5; n=816)	Normal (18.5-25.0; n=49 017)	Overweight (25.0-30.0; n=74 439)	Obese (>30.0; n=44 784)	Female (n=81 091)	Male (n=87 965)

Age started smoking (yrs)	17.3 (4.2)	17.5 (4.8)	17.6 (4.2)	17.3 (4.2)	17.1 (4.3)	17.8 (4.4)	16.9 (4.0)
No of cigarettes smoked per day							
Ever smokers	18.4 (10.1)	16.6 (10.5)	15.9 (8.6)	18.2 (9.6)	21.1 (11.5)	16.1 (8.2)	20.5 (11.2)
Current smokers*	15.8 (8.4)	16.8 (11.1)	15.0 (8.2)	15.6 (8.1)	17.3 (9.0)	14.2 (7.3)	17.4 (9.2)

·N = 37 519 current smokers.

Tables adapted from Carreras-Torres et al., 2018.[46]

Explanation: Information on the distribution of the exposure, outcomes, and other variables helps judge the comparability of groups and generalisability of the findings. Distributions of continuous variables are easily summarised by mean and standard deviation, or by median and percentile range (e.g., 25th and 75th percentiles) in case of asymmetrical distribution. Numbers and percentages best describe categorical variables. Readers can assess group differences better if the descriptive statistics are provided for each category separately. Statistical inference regarding differences between groups should be reserved for the main analysis.[25]

In cohort studies in which the outcome is an event, it is important to report both the number of events and, if appropriate, the event rate, e.g., number of events per person-year. A summary measure of follow-up time, such as mean, median or total follow-up, is also important to understand the period over which events were recorded.

For a time-varying outcome, for which time to event data are available, the summary measures should be presented over time; a figure may help communicate this. In case-control studies, the summary measures are typically presented separately for cases and controls. It may also be helpful to tabulate continuous exposures or outcomes by categories.[25]

10c) If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies.

Example: "Table 2 demonstrates the I^2 test statistic, allowing for an assessment of heterogeneity of the effect of the genetic variants on the outcome across studies." [38]

"Table 2: Genome wide significant single nucleotide polymorphisms (SNPs) for fracture."

Locus	Candidate gene	SNP	Distance to gene (kb)	EA	EAF	Discovery stage*		Replication stage*		Combined*			
						Odds ratio (95% CI)	P	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P	No of fracture cases	I^2
2p16.2	*SPTBN1*	rs4233949	−23.21	G	0.61	1.03 (1.02 to 1.05)	6.9×10^{-5}	1.04 (1.05 to 1.05)	8.9×10^{-11}	1.03 (1.02 to 1.04)	2.8×10^{-14}	185 057	22.4
3p22.1	*CTNNB1*	rs430727	107.2	T	0.45	1.03 (1.02 to 1.05)	1.0×10^{-4}	1.03 (1.02 to 1.04)	1.1×10^{-8}	1.03 (1.02 to 1.04)	5.0×10^{-12}	185 057	0
6q22.33	*RSPO3*	rs10457487	0	C	0.51	1.06 (1.05 to 1.08)	2.3×10^{-15}	1.04 (1.03 to 1.05)	1.7×10^{-15}	1.05 (1.04 to 1.06)	4.8×10^{-28}	185 057	5
6q25.1	*ESR1*	rs2982570	0	C	0.58	1.05 (1.04 to 1.07)	8.1×10^{-12}	1.03 (1.02 to 1.04)	5.2×10^{-10}	1.04 (1.03 to 1.05)	4.5×10^{-19}	185 057	23
7q31.31	*WNT16, CPED1*	rs2908007	−3.25, 24.67	A	0.60	1.08 (1.06 to 1.10)	1.2×10^{-20}	1.05 (1.04 to 1.06)	5.6×10^{-22}	1.06 (1.05 to 1.07)	2.3×10^{-39}	185 055	0
7q21.3	*C7orf76, SHFM1*	rs6465508	0, 0	G	0.34	1.05 (1.03 to 1.07)	4.0×10^{-9}	1.04 (1.03 to 1.05)	4.1×10^{-12}	1.04 (1.03 to 1.05)	2.0×10^{-19}	185 056	35
7p14.1	*STARD3NL*	rs6959212	−89.01	T	0.34	1.04 (1.02 to 1.06)	6.9×10^{-6}	1.02 (1.01 to 1.04)	1.1×10^{-5}	1.03 (1.02 to 1.04)	8.8×10^{-10}	185 057	15.6
7p12.1	*GRB10, COBL*	rs1548607	40.33, −182.4	G	0.32	1.05 (1.03 to 1.07)	3.2×10^{-8}	1.02 (1.01 to 1.04)	2.1×10^{-4}	1.03 (1.02 to 1.05)	4.7×10^{-10}	185 052	40
9q34.11	*FUBP3*	rs7851693	0	G	0.35	1.03 (1.01 to 1.06)	1.3×10^{-4}	1.05 (1.06 to 1.06)	4.8×10^{-16}	1.04 (1.03 to 1.05)	5.0×10^{-19}	185 057	23.5
10q21.1	*MBL2/DKK1*	rs11003047	−90.63	G	0.11	1.09 (1.07 to 1.12)	6.2×10^{-12}	1.08 (1.07 to 1.10)	1.4×10^{-21}	1.09 (1.07 to 1.10)	9.5×10^{-33}	185 057	0
11q13.2	*LRP5*	rs3736228	0	T	0.15	1.05 (1.03 to 1.07)	3.0×10^{-5}	1.07 (1.05 to 1.08)	2.8×10^{-18}	1.06 (1.05 to 1.08)	1.0×10^{-21}	185 056	24.6
14q32.12	*RPS6KA5*	rs1286083	0	T	0.82	1.04 (1.02 to 1.06)	8.8×10^{-5}	1.05 (1.04 to 1.07)	3.0×10^{-14}	1.05 (1.04 to 1.06)	1.6×10^{-17}	185 085	43.3
17q21.31	*SOST, DUSP3, MEOX1*	rs2741856	−4.26, −16.65, 88.02	G	0.92	1.11 (1.08 to 1.14)	2.4×10^{-12}	1.08 (1.06 to 1.11)	5.3×10^{-15}	1.10 (1.07 to 1.11)	3.1×10^{-25}	184 977	0
18p11.21	*FAM210A, RNMT*	rs4635400	0, −7.149	A	0.36	1.06 (1.04 to 1.07)	1.5×10^{-12}	1.03 (1.02 to 1.04)	2.7×10^{-9}	1.04 (1.03 to 1.05)	1.1×10^{-18}	185 057	22
21q22.2	*ETS2*	rs9980072	141.9	G	0.73	1.06 (1.04 to 1.08)	8.4×10^{-12}	1.03 (1.01 to 1.04)	1.8×10^{-5}	1.04 (1.03 to 1.05)	3.4×10^{-13}	185 057	36

EA=effect allele; EAF=effect allele frequency; I^2=index of heterogeneity.

- * Discovery stage (37 857 cases; 227 116 controls); replication stage (147 200 cases; 150 085 controls); combined (185 057 cases; 377 201 controls).

Table reproduced with permission from Trajanoska et al, 2018.[38]

Explanation: Evidence on the consistency of the genetic variant's association with the exposure or outcome helps to understand the degree of heterogeneity of effects. If the estimation is based on a meta-analysis, the number of included studies will also help understand if tests for heterogeneity are properly powered to detect its presence. Presenting 95% confidence intervals along with the I squared statistic is recommended.[86 87]

10d) For *two-sample* Mendelian randomisation:

i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples.

Example: "The genetic variants used for MR were obtained from a GWAS of gallstones conducted in Europeans. A comparison between European and Indian populations with respect to allele frequencies, risk of developing gallstones and gallbladder cancer (GBC) for the genetic variants was made and results are shown in Supplementary Table 1. The allele frequencies between the two populations were generally similar, although with striking differences for some SNPs (e.g. for rs601338, rs1260326, rs174567, rs2469991, rs2290846, where the difference in minor allele frequency was >15%). The risk for developing gallstones and GBC were in broadly the same direction for the SNPs in the Indian population (consistently increased risk for 80% of SNPs in relation to gallstones and 70% SNPs in relation to GBC)." [88]

Explanation: Two-sample Mendelian randomisation analyses assume that the SNP-exposure and SNP-outcome associations are similar in the two samples. For example, this is assumed if the two samples are drawn from the same underlying population. Where this

assumption cannot be made, it should be evaluated by comparing the SNPs' association with the exposure and the outcome in the two samples, whenever the data are available. If the associations are similar in the two samples, heterogeneity in the associations of the SNPs with the exposure and the outcome is less likely to cause bias.

ii. Provide information on the number of individuals who were in both samples for the exposure and for the outcome.

Example: "These genome-wide association study estimates were selected from studies that did not include UK Biobank participants, so as to avoid participant overlap, and therefore, in some cases, the genome-wide association study and subsequent instruments differed from the genome-wide association study studies used for the two-sample mendelian randomisation described previously." [84]

Explanation: If the authors used the same or similar individuals to estimate the SNP-exposure and the SNP-outcome associations, MR-estimates could be biased by a form of the winner's curse.[89] This bias can be overcome by using entirely separate samples to select SNPs and estimate SNP-outcome associations. The bias is a linear function of the number of individuals included in both samples, so the consequences of a small amount of overlap may not be severe.[89]

11. Main results

11a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale.

Example: "The BMI allele score created from the 12 BMI-related SNPs showed a positive dose-response association with BMI (per unit increase 0.14% [0.12%–0.16%], $p=6.30\times10^{-62}$). The BMI allele score was also associated with 25(OH)D concentrations (per unit increase -0.06% [-0.20% to -0.02%], p=0.004)." [37]

Explanation: Reporting the association between the genetic variant and the exposure is required to evaluate the relevance assumption (see item 8b). Comparing levels of exposure across the genotype distribution can also indicate monotonicity and linearity of the genetic effect. It is also useful to report on the association between the genetic variant and the outcome, which can provide an initial indication about the possibility of a causal relationship between the exposure and outcome.

11b) Report causal effect estimates between exposure and outcome, and the measures of uncertainty from the MR analysis, on an intuitive scale, such as odds ratio or relative risk per standard deviation difference.

Example: "In the analysis to establish the direction and causality of BMI-25(OH)D association by the use of the IV ratio, BMI was associated with 25(OH)D: each 10% increase in BMI led to a 4.2% decrease in 25(OH)D concentrations (-7.1% to -1.3%, p=0.005)." [37]

Explanation: If the IV assumptions are not falsified and are generally supported, or sensitivity analyses are robust to violation of the assumptions (item 8b), then causal effect estimates can be reported, preferably on an intuitive scale (e.g. relative risk, risk difference). However, if the homogeneity and monotonicity assumptions are unlikely to hold, it may be preferable not to estimate a causal effect. Instead, results should be interpreted as a test of

whether the exposure has a causal effect on the outcome in at least some individuals during the life-course. In that case, the estimation of a causal effect is replaced by testing for a non-null causal effect.

11c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.

Example: "LDL cholesterol lowering alleles at the *NPC1L1* locus were inversely associated with coronary artery disease (OR for a genetically predicted 1-mmol/L [38.7-mg/dL] reduction in LDL-C of 0.61 [95% CI, 0.42-0.88]; $P = .008$) and directly associated with type 2 diabetes, both individually and collectively (OR for a genetically predicted 1-mmol/L reduction in LDL-C of 2.42 [95% CI, 1.70-3.43], $P < .001$; estimated absolute risk difference, 5.3 incident cases per 1000 person-years for a 1-mmol/L genetically predicted reduction in LDL-C)." [90]

Explanation: In some instances, it may be more clinically meaningful to interpret estimates in terms of absolute risks or risk differences rather than relative risk differences by taking into account baseline risk. A measure of absolute risk can provide an estimate of the excess amount of disease that can be attributed to the exposure over a particular period, which can then be used to estimate the absolute benefit of an intervention aimed at reducing levels of the exposure.

11d) Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure).

Example: "Figure 3: Causal relationships of insomnia symptoms. (**A**) Associations between SNPs associated with frequent insomnia symptoms and CAD. Per-allele associations with risk plotted against per-allele associations with frequent insomnia-symptom risk (vertical and horizontal black lines around points show 95%CI for each polymorphism) are shown for three different MR association tests. (**B**) Forest plot showing the estimates of the effect of genetically increased insomnia risk on CAD. Nearest genes are displayed to the right of the plots. Also shown for each SNP is the 95%CI (gray line segment) of the estimate and the IVW MR, MR-Egger, and weighted-median MR results in red. Sample sizes of each GWAS used in the MR analyses are as follows: frequent insomnia symptoms ($n_{cases} = 129,270$; $n_{controls} = 108,352$), CAD ($n_{cases} = 60,801$; $n_{controls} = 123,504$)." [91]

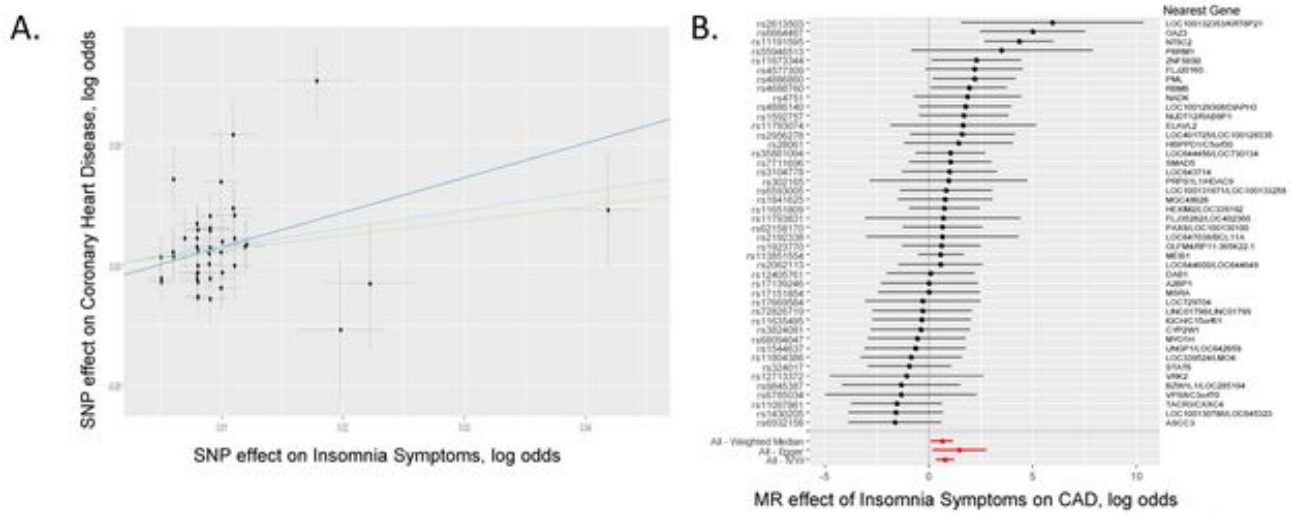


Figure reproduced with permission from Lane et al., 2019.[91]

Explanation: Plots can be useful for examining potential violations of the IV assumptions, especially the exclusion restriction assumption. It is particularly important to report the associations of the exposure and outcome with the genetic variants individually, which can be presented using a scatter or funnel plot.[50] The scatter plot depicts the relationship of the genetic effects on the exposure versus the genetic effects on the outcome, with the slope of the line corresponding to the estimated causal effect, with an intercept that is fixed at the origin (except for the case of MR Egger regression, see item 8b). A funnel plot, in which causal effect estimates for variants are plotted against their precisions, can be used to perform a visual inspection for asymmetry, which may be indicative of horizontal pleiotropy.[50] Forest plots, which plot the causal estimate obtained from each genetic variant allow for a visual inspection of heterogeneity around the overall causal estimate.[79]

12. Assessment of assumptions

12a) Report the assessment of the validity of the assumptions.

Examples:

Relevance assumption

"The myopia allele score explained 4.32% (*F*=3155) of the variance in average mean spherical equivalent refractive error of participants in UK Biobank and the education allele score explained 0.71% (*F*=464) of the variance in time spent in education. We selected these genetic variants to use as instrumental variables because of their robust association with time spent in education and myopia, allowing us to construct strong aggregate instrumental variables for making mendelian randomisation inferences. The large F statistics suggested that these analyses would not be affected by weak instrument bias." [82]

Independence assumption

"In tests of the association between the allele scores for time spent in education and myopia with potential confounders, there was evidence that the geographical coordinate, northing (measured northward distance in UK) was negatively associated with time spent in education (β=−1.6e-6, 95% confidence interval −1.8e-6 to −1.5e-6) and positively with

refractive error (β=1.2e-6, 9.8e-7 to 1.3e-6). Northing was also associated with the time spent in education (P=7e-5) and myopia (P=6e-3) allele scores (see supplementary table 2). Compared with standard regression, the confounding bias plot suggested that inclusion of the northing variable in the instrumental variable analysis would result in a greater degree of bias for the education allele score but not for the myopia allele score." [82]

Exclusion restriction

"MR-Egger, weighted mode, and weighted median methods […] yielded similar causal estimates in magnitude and direction, such that increasing time spent in education led to a more myopic refractive error (by −0.17 to −0.40 dioptres/y), whereas there was little evidence that a more myopic refractive error led to more time spent in education […] There was little evidence that the Egger intercept deviated from zero either for more time in education causing refractive error (intercept=0.007, SE=0.006, P=0.2) or refractive error causing more time in education (intercept=−0.002, SE=0.007, P=0.8), indicating that there was little evidence for directional genetic pleiotropy." [82]

Homogeneity

"We observed a J shaped relation between genetically predicted BMI and all-cause mortality. The curved shape of the relation was more pronounced in UK Biobank—with higher risk both in underweight participants and in overweight or obese participants. The lowest risk for the overall population was at a BMI of around 22-23 in the HUNT Study and around 25 in UK Biobank." [92]

Explanation: Authors should report the results from assessing the validity of the IV assumptions, as described under item 7 and **Box 4**. These examples illustrate assessments of these assumptions, but do not represent an exhaustive list of possible assessments or assumptions.

12b) Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value).

Example: "Cochran's Q and I^2 statistics were calculated to check for the presence of heterogeneity (dispersion of SNP effects) which can indicate pleiotropy. We found little evidence of heterogeneity for the association between body mass index and wellbeing (see supplementary table S8 for further information)." [93]

Example: "Chen et al. used a single variant in the ALDH2 gene to study the effects of alcohol intake on risk of hypertension. Among males, the variant-hypertension association was an odds ratio of 2.42. The E-value then is 4.27. The E-value for the lower limit of the confidence interval (1.66) is 2.71. As the analysis was conducted in an ethnically homogeneous Asian population, this E-value may be large enough to reasonably conclude that any residual confounding by ancestry is unlikely to explain away the effect." [64]

Explanation: Cochran's Q and I^2 statistics can be used to assess evidence of heterogeneity of causal effects estimated by each of the genetic variants.[94] Evidence of heterogeneity suggests that there is at least one proposed instrument for which at least one of the IV assumptions fails to hold. The E-value[64] can be used to understand the degree to which unmeasured confounding may explain findings. A large E-value may help support that confounding by ancestry is unlikely to explain a non-null effect.

13. Sensitivity analyses and additional analyses

13a) Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions.

Example: "The fixed-effect inverse-variance weighted and Egger regression estimates suggest an inverse causal effect of CRP on CAD risk (Table1). However, the corresponding random-effects analyses imply that there is no convincing evidence for a causal effect. Moreover, the simple median estimate is in the opposite direction. This arises because, although the strongest genetic variants have negative causal estimates, the majority of genetic variants have positive causal estimates. The inconsistency of the estimates from different methods indicates that the genome-wide significant variants for CRP are not all valid instrumental variables, and that a causal conclusion based on these variants would be unreliable." [69]

Explanation: Authors should report on, and compare, results obtained from different approaches used to assess the robustness of conclusions to violation of the IV assumptions, as described in **Section 7** and **Box 4**. If the results from all the approaches are largely consistent, there can be more confidence in drawing robust conclusions regarding the presence and magnitude of a causal effect.

13b) Report results from other sensitivity analyses or additional analyses.

Examples:

Independent replication

"The associations of genetically predicted body mass index and waist circumference with risk of being a smoker were replicated in the TAG data (1.19 (1.06 to 1.33) and 1.32 (1.15 to 1.52), respectively." [46]

Validation of instruments

"The MR-PRESSO method identified one outlier SNP for heart failure, six outlier SNPs for coronary artery disease, and 11 outlier SNPs for arterial hypertension. Outlier-correction did not materially change the OR estimates for heart failure (1.13; 95% CI 1.08–1.17), coronary artery disease (1.08; 95% CI 1.06–1.10), or arterial hypertension (1.10; 95% CI 1.08–1.12). No outlier SNPs were identified in the MR-PRESSO analysis of the other outcomes." [95]

Simulations

"Figure 2 shows that TSLS is positively biased when there is positive cross-trait assortative mating on *X* and *Y*. The bias increased proportionally with increasing the degree of assortment. However, both TSLS (2) (i.e., adjusting for parent's allele scores) and TSLS (3) (i.e., jointly modelling individual's and parental effects, using nontransmitted allele scores as instruments of parental phenotype) were unbiased with false discovery rates close to 5%." [57]

Explanation: Results of sensitivity analyses or additional analyses, such as independent replication, validation of instruments, and simulation studies, should be presented if they have been performed, as described under item 8.

13c) Report any assessment of direction of causality (e.g., bidirectional MR).

Example: "The BMI allele score was also associated with 25(OH)D concentrations (per unit increase −0.06%, [−0.10% to −0.02%], p=0.004) while no association with BMI was seen for either the vitamin D synthesis or metabolism allele scores (per allele in synthesis score: 0.01% [−0.17% to 0.20%], p=0.88, metabolism allele score: 0.17% [−0.02% to 0.35%], p=0.08])." [37]

Explanation: Bidirectional MR can be used to orient the causal direction(s) of effect. This is done using two independent sets of genetic variants related to the exposure and outcome separately, and performing MR analyses to appraise causality in both directions.[96]

13d) When relevant, report and compare with estimates from non-MR analyses.

Example: "Using the Durbin-Wu-Hausman test for endogeneity, we found weak evidence that the instrumental variable estimate using the time spent in education allele score differed from the observational point estimate (Durbin-Wu-Hausman P=0.06), with the instrumental variable estimate suggesting a larger negative association." [82]

Explanation: It is relevant to describe important differences between MR estimates and estimates from non-MR analyses. Each study design has different types of biases and may have different degrees of statistical power. Putting the MR results in context will help the reader understand if the strengths and weaknesses of MR allow for results that support or contradict prior evidence. In general, causal inference can be presented in a triangulation framework, evaluating the overall body of evidence from several different approaches.[97] [98]

13e) Consider additional plots to visualize results (e.g., leave-one-out analyses).

Example: "Figure S3. Leave-one-out analysis: each row represents a two-sample MR analysis of BMI on subjective wellbeing using all of the genome-wide significant SNPs available from Locke et al. except for the SNP listed on the y-axis. The point represents the effect size with that SNP removed and the line represents the standard error.

Leave-one-out analysis was conducted using MR Base to identify if any individual SNPs were driving the association between BMI and wellbeing. […] The SNP with the largest contribution to the effect is rs1421085 located on chromosome 16 in the second intron of the FTO (fat mass and obesity associated) gene. FTO has been repeatedly associated with obesity in different populations. However, the biological consequences of intronic FTO SNPs are still unknown. They are currently thought to play a regulatory role in FTO gene expression in the hypothalamus. Although research is not completely certain of the role of FTO, its large effect size and robust association with obesity suggest that this gene has the largest effect in the two-sample MR because of its BMI effect size rather than because of pleiotropic effects." [93]

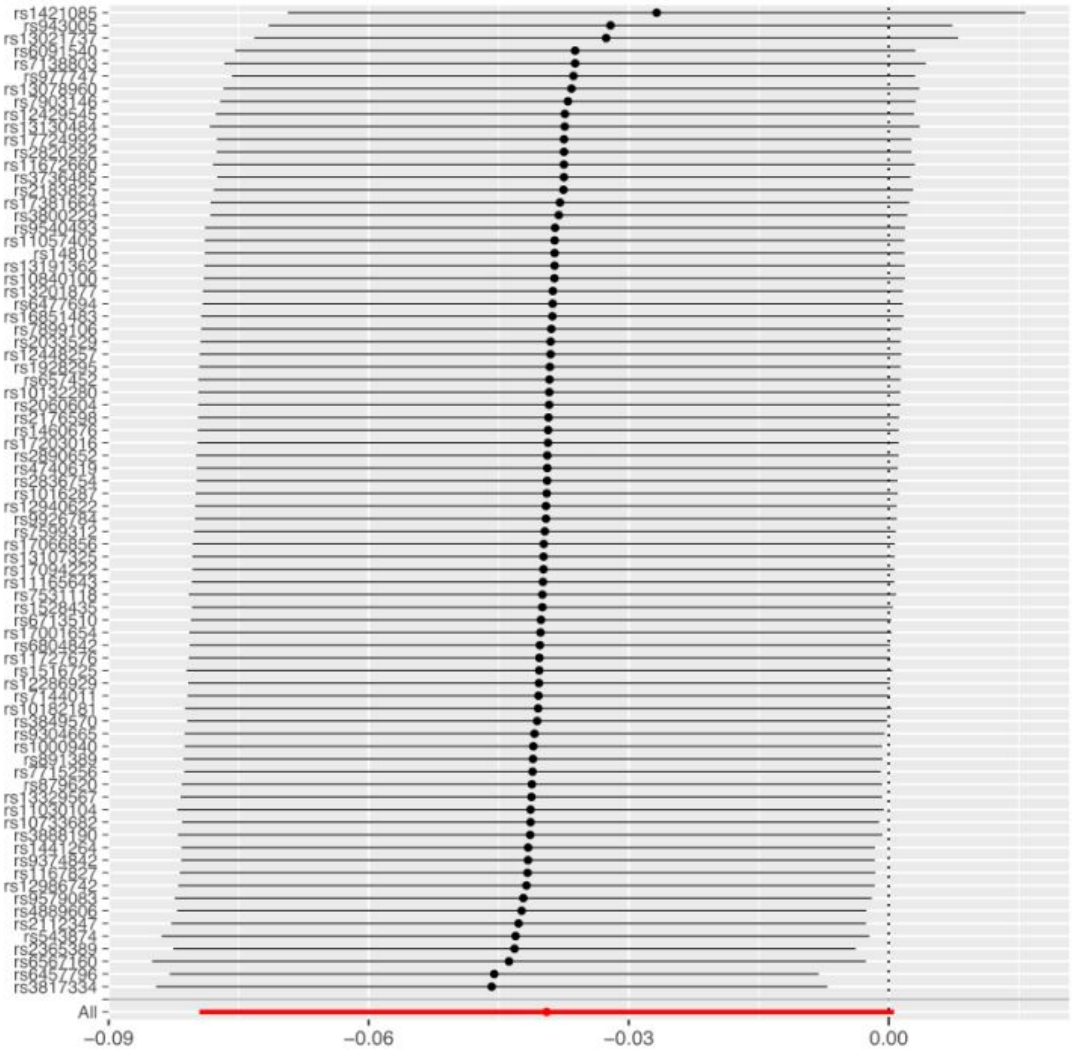


Figure reproduced with permission from Wootton et al., 2018.[93]

Explanation: Additional plots may also aid in the visualisation of results, assessing assumption violation, and detecting potential influential/outlier points. These include the leave-one-out plot,[79] Radial plot[99] and plots of each genetic variant against their studentized residuals or Cook's distance) for outlier assessment.[31]

DISCUSSION

The discussion should address the important issues pertaining to study interpretation and validity.[100] Structured discussions can help authors avoid over-interpreting results, and act as a guide for readers of the article.[101] [102]

14. Key results

Summarize key results with reference to study objectives.

Example: "Based on comprehensive genetic data from nearly 450 000 individuals, our study provides evidence that differences in body mass index and body fat distribution causally influence different aspects of smoking behaviour, including the risk of individuals taking up smoking, smoking intensity, and smoking cessation. These results highlight the role of obesity in influencing smoking initiation and cessation, which could have implications for public health interventions aiming to reduce the relevance of these important risk factors." [46]

Explanation: The discussion should begin with a summary of the main results and a statement of their importance. This section reminds readers of the study questions and its primary findings. Further, it helps readers assess whether the interpretations that follow are consistent with the results. It is good practice to keep the summary in the perspective of the main study objectives and focus on the pre-specified hypothesis, reporting the estimates of the investigated causal relationship in the given population.[101]

15. Limitations

Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and your efforts to address them.

Example: "As in any Mendelian randomisation analysis, several assumptions were made, including that the genetic instruments were associated with the risk factor of interest, were independent of potential confounders, and could only affect the outcome through the risk factor and not through alternative pathways (that is, through pleiotropy). We note that the first assumption was satisfied because robustly associated gene variants were identified from the largest genome wide association study for each obesity parameter. Whether the other two assumptions held was not readily testable, although we conducted thorough sensitivity analyses that did not highlight any obvious violation of these assumptions. Secondly, a potential confounder of our results was population stratification by sociodemographic factors. Indeed, it was previously shown that the genetic instrument for body mass index was associated with various factors related to social class among women, including lower annual household income and level of deprivation. However, no such associations were seen in men. In our study, the associations between the genetic instruments of obesity and individuals taking up smoking and smoking intensity were consistently observed in both men and women, separately, and also when we excluded SNPs that were potentially linked to social deprivation. Therefore, apart from the inverse association between body fat percentage and smoking cessation observed in women only, population stratification by sociodemographic factors would not seem likely to explain those results." [46] ***WA***

Explanation: Authors should address the plausibility of all the IV assumptions. This is especially important as many of the assumptions are not empirically verifiable. Authors could consider, for example, the possibility that (residual) genotype-phenotype confounders (such as population structure, genetic nurture, or assortative mating) could lead to a violation of the independence assumption. When evaluating a potential violation, authors should identify the sources of a violation that could affect results and discuss the relative importance of different violations, including the likely direction and magnitude of any bias they could induce.

It is also important that authors discuss the precision of the results. Imprecision can be due to several features of the study design. For example, an IV estimate's precision from a meta-analysis of multiple SNPs will usually be greater than that for a single SNP. Suppose SNPs are chosen based on meeting a p-value criterion in a discovery GWAS. In that case, authors should consider factors that affect this GWAS's power to detect SNPs, such as sample size and measurement error. IV estimates will also be more precise when estimated from larger datasets, because the standard error for the SNP effect estimates, used to calculate IV estimate, will be smaller.

16. Interpretation

16a) Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison to other studies.

Example: "These Mendelian randomization analyses suggest that the causal effect of CETP (*cholesteryl ester transfer protein*) inhibition on the risk of cardiovascular events appears to be determined by changes in the concentration of apoB-containing lipoproteins rather than changes in LDL-C or HDL-C level." [103]

Explanation: Provide a cautious interpretation of the overall results. When comparing with results from other studies, consider how the results may differ from previous estimates and discuss possible reasons for these differences. Such reasons could include violations of IV assumptions, imprecision, different estimation methods and different studied populations. Consider that the overall results should be interpreted in the context of other studies that assessed the study question using different designs, allowing for triangulation of results (see item 13d). When interpreting the effect size, discuss assumptions underlying any extrapolations of effect size and how they may have influenced results.

16b) Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable.

Example: "The association between pubertal timing and weight status is complex and plausibly bidirectional. Increased adiposity in childhood has been linked to earlier pubertal maturation, although this relationship may be nonlinear in boys. Furthermore, several studies report evidence for an association between earlier age at puberty and later obesity. Therefore, we sought to control for both genetically predicted adult and childhood BMI, and we observed a similar magnitude of attenuation in the association between pubertal timing and risk of MS. However, there is a strong association between childhood and adult BMI, which limits the exploration of age-specific effects. Nonetheless, postpubertal rather than childhood obesity is most clearly related to MS susceptibility, making the association

between pubertal timing and adult obesity the most likely mediator of the effect of age at puberty on risk of MS. Because it appears that BMI and pubertal timing are in the same causal biological pathway, the association of the selected genetic variants with both exposures represents an example of vertical pleiotropy due to shared biological underpinnings and thus does not bias the MR estimates." [34]

Explanation: While the biological mechanisms that allow genetic variants to be used as instrumental variables are often unknown, the discussion should consider possibilities. Doing so will enable the reader to put the MR results in context about possible biological mechanisms, allowing for a better understanding of the plausibility of causal relationships.

16c) Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions.

Example: "Although uncertainty remains around the precise function of each of the 162 SNPs, their degree of pleiotropy with cardiac traits, and the mechanisms by which these genetic variants exert their cardioprotective influence, conclusions can still be drawn. […] we note that interventions should be accompanied by careful monitoring for unforeseen side effects, especially in those people who may not thrive when forced into extended educational settings, which may otherwise aggravate health inequalities." [104]

Explanation: Investigators should describe the potential impact of the results on clinical practice or public policy, if any. Since many interventions cannot be tested in randomized clinical trials, MR evidence may help to better understand the possible causal effect of the exposure on the outcome. Such statements should be made with caution and in light of evidence from other sources, such as other observational and experimental studies when available. Since clinical and policy interventions may have different effect sizes than the genetic variants included in the MR study, extrapolation of this evidence should be clearly described and cautious.

Box 5. Interpretation of causal effect estimates

Various considerations should be taken into account when interpreting causal estimates. If the homogeneity assumption is plausible, along with the other assumptions (see **Box 3**), then the causal estimate will represent the average causal effect of the exposure on the outcome in the studied population. If the homogeneity assumption cannot be made, but the monotonicity assumption is plausible, then the causal estimate can be used to represent the local average treatment effect.[81] Caution is especially warranted when interpreting effect estimates with a binary exposure.[105] In this instance, the homogeneity and monotonicity assumptions are less likely to be plausible. Also, if the exposure is a dichotomization of a continuous risk factor, this poses a further threat to the violation of the exclusion restriction assumption.[105] An additional consideration, which is particularly pertinent to the two-sample MR setting, is whether the causal effect can truly be attributed to the binary exposure. For example, when two-sample MR studies are carried out in exposure samples that contain only a small number of participants who have experienced the exposure in question, it would be misleading to interpret the effects as being those of the exposure itself. Rather the causal effect estimate should be interpreted as reflecting the effects of the genetic liability to the exposure.[106] Finally, an important component of interpretation is defining the time period.[55] Usually MR studies are interpreted as a "lifetime" effect of the exposure, but some settings (e.g. MR studies in pregnant women to study prenatal exposures) lend themselves to studies of period effects.

17. Generalizability

Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure.

Example: "Our Mendelian randomisation work examined a linear relation between vitamin D levels and fracture risk. We did not test for the possibility of a threshold dependent relation—that is, effects that could be present only at very low levels of vitamin D. […] Finally, the non-significant trend observed for vitamin D towards having increased risk of fracture could be attributed to the selection of healthy people (that is, participants with very low levels of vitamin D and fracture, as well as those who are older, frail, and physically impaired, could have been under-represented in the studies included in the GWAS meta-analyses). Therefore, the vitamin D estimates of the current study cannot be generalised to these groups of older people." [38]

Explanation: The generalisability of a study is the extent to which the study's results apply to circumstances different from the ones in which the study was conducted.[107] For example, findings from a cohort of a specific age group collected in the past may not apply to people currently in the same age strata.[108]

MR studies can fail to generalise in other ways. For example, because genetic variants may not have a constant effect over the entire life course, it is important to consider if the effect estimate derived in the study would generalise to other exposure periods. For example, if the effect of the exposure on the outcome is time-dependent, or only occurs during a critical period, MR estimates may be misleading if used to guide future interventions if they occur outside of this time frame. Likewise, if the effect of an exposure on an outcome is cumulative over many years, MR can overestimate the effect when compared to short term interventions.[55 109]

Also, MR estimates are directly calculated only for the exposure range caused by differences in allele(s). Applying MR results, therefore, may not generalise to a wider exposure range. Further, if the MR estimate was derived from a population subgroup, it may not be generalizable beyond that subgroup.

OTHER INFORMATION

18. Funding

Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based.

Example: "Funding: The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie, de la Science et de l'Innovation du Québec through Genome Québec and grant PSR-SIIRI-701, the National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710), and the European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al.25 RCR, ELA, BMB, CLR,

RMM, MM, DAL, and GDS are members of the MRC Integrative Epidemiology Unit at the University of Bristol funded by the Medical Research Council (grant Nos MM_UU_00011/1, MC_UU_00011/2, MC_UU_00011/5, MC_UU_00011/6, and MC_UU_00011/7). RCR is a de Pass VC research fellow at the University of Bristol. This study was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, National Institute for Health Research, or Department of Health and Social Care. This work was also supported by Cancer Research UK (grant No C18281/A19169) and the Economic and Social Research Council (grant No ES/N000498/1). SEJ is funded by the Medical Research Council (grant No MR/M005070/1). TMF is supported by the European Research Council (grant No 323195:GLUCOSEGENES-FP7-IDEAS-ERC). MNW is supported by the Wellcome Trust Institutional Strategic Support Award (grant No WT097835MF).” [110]

Explanation: The source of research funding can lead to bias or perceptions of bias in the design, conduct, or interpretation of research.[111] [112] This is of special concern when research is funded by an entity that has an interest in outcomes that are favourable to its own commercial, academic, or other interests.[113] Authors should disclose all funding sources and provide detailed information about the role of funders in developing the research question, collecting data, analysing data, selecting investigators, reviewing results, preparing the manuscript, or approving the manuscript for submission or publication. Other sources of influence may include employers, political appointees, and government researchers. Describing the source of funding allows readers to evaluate the work's credibility and trustworthiness in light of any potential influence from funders. Authors should disclose funding sources for biobanks or other repositories or databases used in their study because these entities also have commercial interests that may influence research integrity.[114] [115]

19. Data and data sharing.

Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where.

Example: “Data sharing: The data reported in this paper are available by application directly to the UK Biobank. The genetic associations with the outcomes in the UK Biobank and CARDIoGRAMplusC4D consortium are provided in the supplementary data. Software code in R for implementing the mendelian randomisation analysis, including the principal components analysis, is provided in the supplementary note.” [116]

Explanation: Original data are needed by readers or researchers who wish to evaluate or replicate analyses. Many funders and journals encourage or require data sharing, and provide guidance to authors about the content of any explicit data sharing statement required by their journal. Consensus is building that data sharing is “an inseparable part of the research process”.[117] Ideally, a data sharing plan should be developed when a study is being organized, and described in the study protocol and journal publications. The plan and any subsequent data sharing statement in the article should indicate, at a minimum, what data are available (e.g. individual participant data, statistical analysis plan, documents

related to the study, biobank or other database information) and how the data can be accessed. Contact information for the person or organization holding the data and a description of the mechanism that will be used to share data should be provided. The statement should also describe any time limits on data availability, processes, and standards applied to requests for data (such as requirements for a research protocol or review of applications by a review board) and, if known, whether there is a charge to obtain the data. When data come from multiple sources, and different conditions apply, consider the use of a table format instead of a text statement.

20. Conflicts of Interest. All authors should declare all potential conflicts of interest.

Example: "Competing interests: All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form at www.icmje.org/coi_disclosure.pdf. ARC, DG, TT, JV, REW, GH, RM, SS, SB, GDS, MVH, IT, and AD declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. MRM reports grants from Pfizer and non-financial support from GlaxoSmithKline, outside the submitted work. NMD reports grants from ESRC, grants from MRC, during the conduct of the study; grants from GRAND/Pfizer for unrelated research, outside the submitted work. AET reports grants from Pfizer, outside the submitted work. LDH reports grants from MRC, during the conduct of the study. DW reports grants from NIH, during the conduct of the study." [84]

Explanation: Financial connections between researchers and commercial or other entities and firmly held ideological or intellectual views can lead to bias in the design, conduct, or reporting of study results. When such interests are not disclosed, public trust in the research enterprise is eroded.[118] According to the ICMJE, "conflict of interest exists when professional judgment concerning a primary interest (such as patients' welfare or the validity of research) may be influenced by a secondary interest (such as financial gain). Perceptions of conflict of interest are as important as actual conflicts of interest".[119] Authors should err on the side of disclosing all matters that might be considered relevant by readers.

Conclusions

The STROBE-MR reporting guideline proposes a minimum set of items supporting authors to clearly communicate what was planned, what was done, and what was found in an MR study. Similar to the STROBE guidelines[24 25] for the classical epidemiological study designs – cohort, case-control and cross-sectional studies – the goal is not to be prescriptive of study conduct or limit creativity in the field. Rather, the STROBE-MR guideline is intended to facilitate clear and comprehensive reporting to enable an appraisal of a study's quality, limitations, and generalisability of findings. The checklist is not intended as a formal tool for assessing the methodological or reporting quality of MR studies, and should not be transformed into a quality scale.[120 121] STROBE-MR should also not be seen as a formal guideline to design and conduct MR studies. However, some items and text might be useful when designing or conducting an MR study, and this E&E document may be useful to inform methodological decisions, particularly for researchers with less experience in MR research.

We invite readers to comment on STROBE-MR and suggest improvements to the checklist, explanations, and examples. Checklist and E&E document are living documents that we intend to keep up to date on a dedicated website (http://www.strobe-mr.org/). We encourage journals to endorse these guidelines using clear language regarding what they expect from authors and include this information in their Instructions to Authors. For example, journals could ask authors to submit completed checklists and peer reviewers to use them as part of their review.[27] The STROBE-MR guidance will be included in the EQUATOR Network website (www.equator-network.org), which provides a comprehensive collection of reporting guidelines and other resources.[122] In addition, we welcome and wish to be involved in initiatives to translate the checklist and E&E document to other languages.

Article Information

Acknowledgements

We are grateful to everyone who commented on previous versions of the STROBE-MR checklist. Particular thanks to Jeremy Labrecque, Philip Haycock and Ryan Au Yeung for their helpful suggestions. We are grateful to the Medical Research Council Integrative Epidemiology Unit within the Bristol Medical School, University of Bristol, UK, for hosting the May 2019 workshop.

Contributors to the STROBE-MR Initiative

The following persons have contributed to the content and elaboration of the STROBE-MR checklist: George Davey Smith, Neil M Davies, Niki Dimou, Matthias Egger, Valentina Gallo, Robert M Golub, Julian PT Higgins, Claudia Langenberg, Elizabeth W Loder, J Brent Richards, Rebecca C Richmond, Veronika W Skrivankova, Sonja A Swanson, Nicholas J Timpson, Anne Tybjaerg-Hansen, Tyler J VanderWeele, Benjamin AR Woolf, James Yarmolinsky.

Author contributions

All authors contributed to the writing of the article and approved of its final version. VWS, RCR and BARW prepared the first draft of the checklist and discussion material for the workshop. VWS and JBR took care of practical coordination of the STROBE-MR. ME and GDS initiated STROBE-MR and organised the workshop; ME obtained the funding. ME, GDS and JBR oversaw the project.

Funding

All STROBE-MR Initiative members are volunteers. Support for this initiative was provided by the International Swiss National Science Foundation (SNSF), NIHR Biomedical Research Centre at University Hospitals Bristol, Weston NHS Foundation Trust, and the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the SNSF, NIHR, Weston NHS FT or the University of Bristol. Funders had no role in the design, conduct or results interpretation in this project.

Individual authors funding: The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_00011/1]. **RCR** is a de Pass Vice Chancellor's Research Fellow at the University of Bristol. **NMD** is supported by a Norwegian Research Council (grant number 295989). **SAS** was supported by a NWO/ZonMW Veni grant (91617066). **NJT** is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit (MC_UU_00011) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). **JPTH** is supported by NIHR Bristol Biomedical

Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. JPTH is a member of the MRC Integrative Epidemiology Unit at the University of Bristol. JPTH is a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0617-10145). The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR, MRC, or the Department of Health and Social Care. **ME** is supported by the Swiss National Science Foundation (grant number 189498). **JBR** is supported by the Canadian Institutes of Health Research (CIHR: 365825; 409511), the Lady Davis Institute of the Jewish General Hospital, the Canadian Foundation for Innovation, the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada and the Fonds de Recherche Québec Santé (FRQS). Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. These funding agencies had no role in the design, implementation or interpretation of this study. **TJVW** was funded by NIH grant R01 CA222147. **BARW is** funded by the Economic and Social Research Council (ESRC) South West Doctoral Training Partnership (SWDTP) 1+3 PhD Studentship Award (ES/P000630/1)

Competing Interests

We have read and understood the BMJ Group policy on declaration of interests and declare the following interests: Elizabeth Loder—*BMJ* clinical epidemiology editor—played no part in the peer review or decision making of this paper at the editorial level, and contributed solely as an author.

Transparency declaration

Dr Richards affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. BMJ [Internet]. 2002;325(7378):1437–8. Available from: https://pubmed.ncbi.nlm.nih.gov/12493654

2. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? Lancet (London, England). 2004 May;363(9422):1724–7.

3. Taubes G. Epidemiology faces its limits. Vol. 269, Science (New York, N.Y.). United States; 1995. p. 164–9.

4. Sattar N, Preiss D. Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? Vol. 135, Circulation. United States; 2017. p. 2369–72.

5. Ortolá R, García-Esquinas E, Soler-Vila H, Ordovas JM, López-García E, Rodríguez-Artalejo F. Changes in health status predict changes in alcohol consumption in older adults: the Seniors-ENRICA cohort. J Epidemiol Community Health. 2019 Feb;73(2):123–9.

6. Dekkers OM, Vandenbroucke JP, Cevallos M, Renehan AG, Altman DG, Egger M. COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. PLOS Med. 2019 Feb;16(2):e1002742.

7. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000 Aug;29(4):722–9.

8. Davey Smith G. Capitalizing on Mendelian randomization to assess the effects of treatments. J R Soc Med [Internet]. 2007;100(9):432–5. Available from: https://pubmed.ncbi.nlm.nih.gov/17766918

9. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. Hum Mol Genet. 2019;28(R2):R170–9.

10. Holmes M V, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ Br Med J. 2014 Jul;349:g4164.

11. Lawlor DA, Wade K, Borges MC, Palmer TM, Hartwig FP, Hemani G, et al. A Mendelian Randomization dictionary: Useful definitions and descriptions for undertaking, understanding and interpreting Mendelian Randomization studies. OSF Prepr. 2019;

12. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant. 1991;7 Suppl 3:9–12.

13. Davey Smith G. Capitalising on Mendelian randomization to assess the effects of treatments. [Internet]. The James Lind Library. 2006. Available from: https://www.jameslindlibrary.org/articles/capitalising-on-mendelian-randomization-to-assess-the-effects-of-treatments/

14. Keating S, de Witte T, Suciu S, Willemze R, Hayat M, Labar B, et al. The influence of HLA-matched sibling donor availability on treatment outcome for patients with AML: an analysis of the AML 8A study of the EORTC Leukaemia Cooperative Group and GIMEMA. European Organization for Research and Treatment of Cancer. Gruppo Ita. Br J Haematol. 1998;102(5):1344–53.

15. Bleakley M, Shaw PJ, Nielsen JM. Allogeneic bone marrow transplantation for childhood relapsed acute lymphoblastic leukemia: comparison of outcome in patients with and without a matched family donor. Bone Marrow Transplant. 2002;30(1):1–7.

16. Balduzzi A, Valsecchi MG, Uderzo C, De Lorenzo P, Klingebiel T, Peters C, et al. Chemotherapy versus allogeneic transplantation for very-high-risk childhood acute lymphoblastic leukaemia in first complete remission: comparison by genetic randomisation in an international prospective study. Lancet (London, England). 2005;366(9486):635–42.

17. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. Int J Epidemiol. 2003;32(1):1–22.

18. Davey Smith G, Ebrahim S. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. In: Weinstein M, Vaupel JW, Wachter KWBT-BS, Weinstein M, Vaupel JW, Wachter KW, editors. Biosocial Surveys [Internet]. Washington (DC); 2008. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: http://www.ncbi.nlm.nih.gov/books/NBK62436/

19. Davey Smith G. Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene × Environment Interactions. Perspect Psychol Sci. 2010;5(5):527–45.

20. Attermann J, Obel C, Bilenberg N, Nordenbæk CM, Skytthe A, Olsen J. Traits of ADHD and autism in girls with a twin brother: a Mendelian randomization study. Eur Child Adolesc Psychiatry. 2012;21(9):503–9.

21. Boef AGC, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. Int J Epidemiol. 2015;44(2):496–511.

22. Lor GCY, Risch HA, Fung WT, Au Yeung SL, Wong IOL, Zheng W, et al. Reporting and guidelines for mendelian randomization analysis: A systematic review of oncological studies. Cancer Epidemiol. 2019;62:101577.

23. Diemer EW, Labrecque JA, Neumann A, Tiemeier H, Swanson SA. Mendelian randomisation approaches to the study of prenatal exposures: A systematic review. Paediatr Perinat Epidemiol. 2020;n/a(n/a).

24. Elm von E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806 LP – 808.

25. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLoS Med. 2007;4(10):e297.

26. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) Statement. JAMA [in progress]. 2021;

27. Moher D, Schulz KF, Simera I, Altman DG. Guidance for Developers of Health Research Reporting Guidelines. PLOS Med. 2010;7(2):e1000217.

28. Davey Smith G, Davies NM, Dimou N, Egger M, Gallo V, Golub R, et al. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies

[Internet]. 2019 Dec. Available from: https://peerj.com/preprints/27857

29. MR Conference 2019. MRC Integrative Epidemiology Unit. University of Bristol [Internet]. University of Bristol; 2020. Available from: https://www.bristol.ac.uk/integrative-epidemiology/seminars/2019/mr-conference-2019.html

30. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2020;4(186).

31. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Davey Smith G, et al. BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. Diabetes. 2016;65(10):3002.

32. Ligthart S, Vaez A, Võsa U, Stathopoulou MG, de Vries PS, Prins BP, et al. Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. Am J Hum Genet. 2018;103(5):691–706.

33. Burgess S, Ference BA, Staley JR, Freitag DF, Mason AM, Nielsen SF, et al. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. JAMA Cardiol. 2018;3(7):619–27.

34. Harroud A, Morris JA, Forgetta V, Mitchell R, Davey Smith G, Sawcer SJ, et al. Effect of age at puberty on risk of multiple sclerosis. Neurology. 2019;92(16):e1803.

35. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020;52(10):1122–31.

36. Larsson SC, Burgess S, Michaëlsson K. Association of Genetic Variants Related to Serum Calcium Levels With Coronary Artery Disease and Myocardial Infarction. JAMA. 2017;318(4):371–80.

37. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLOS Med. 2013;10(2):e1001383–e1001383.

38. Trajanoska K, Morris JA, Oei L, Zheng H-F, Evans DM, Kiel DP, et al. Assessment of the genetic and clinical determinants of fracture risk: genome wide association and mendelian randomisation study. BMJ. 2018;362:k3225.

39. Tyrrell J, Jones SE, Beaumont R, Astley CM, Lovell R, Yaghootkar H, et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. BMJ. 2016;352:i582.

40. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, et al. An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet. 2019;51(2):258–66.

41. Johansson M, Carreras-Torres R, Scelo G, Purdue MP, Mariosa D, Muller DC, et al. The influence of obesity-related factors in the etiology of renal cell carcinoma—A mendelian randomization study. PLOS Med. 2019;16(1):e1002724.

42. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30(7):543–52.

43. Lotta LA, Wittemans LBL, Zuber V, Stewart ID, Sharp SJ, Luan J, et al. Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. JAMA. 2018;320(24):2553–63.

44. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al. Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. JAMA. 2016;315(11):1129–40.

45. Association WM. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2013;310(20):2191–4.

46. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ. 2018;361(NA):k1767.

47. Mokry LE, Ross S, Timpson NJ, Sawcer S, Davey Smith G, Richards JB. Obesity and Multiple Sclerosis: A Mendelian Randomization Study. PLOS Med. 2016;13(6):e1002053.

48. Swanson SA, Hernán MA, Miller M, Robins JM, Richardson TS. Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes. J Am Stat Assoc. 2018;113(522):933–47.

49. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996;91(434):444–55.

50. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.

51. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304–14.

52. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.

53. Davies NM, Holmes VM, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;(362):k601.

54. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology. 2014;25(3):427–35.

55. Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials. Epidemiology. 2017;28(5):653–9.

56. Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. Int J Epidemiol. 2018;47(4):1289–97.

57. Hartwig FP, Davies NM, Davey Smith G. Bias in Mendelian randomization due to assortative mating. Genet Epidemiol. 2018;42(7):608–20.

58. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun. 2020;11(1):3519.

59. Swanson SA. A Practical Guide to Selection Bias in Instrumental Variable Analyses. Epidemiology. 2019;30(3):345–9.

60. Bound J, Jaeger DA, Baker RM. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. J Am Stat Assoc. 1995;90(430):443–50.

61. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control outcomes in Mendelian Randomisation to detect potential population stratification or selection bias. bioRxiv. 2020;2020.06.01.128264.

62. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010;19(6):537–54.

63. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. Epidemiology. 2015;26(4):498–504.

64. Swanson SA, VanderWeele TJ. E-Values for Mendelian Randomization. Epidemiology. 2020;31(3):e23–4.

65. Davies NM, Thomas KH, Taylor AE, Taylor GMJ, Martin RM, Munafò MR, et al. How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. Int J Epidemiol. 2017;46(6):2067–77.

66. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60(7):578 LP – 586.

67. Mills HL, Higgins JPT, Morris RW, Kessler D, Heron J, Wiles N, et al. Detecting heterogeneity of intervention effects using analysis and meta-analysis of differences in variance between arms of a trial. medRxiv. 2020;2020.03.07.20032516.

68. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions. Am J Epidemiol. 2012;175(4):332–9.

69. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30–42.

70. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268–74.

71. Conley T, Hansen C, Rossi P. Plausibly Exogenous. Rev Econ Stat [Internet]. 2012;94(1):260–72. Available from: https://econpapers.repec.org/RePEc:tpr:restat:v:94:y:2012:i:1:p:260-272

72. Labrecque JA, Swanson SA. Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures. Am J Epidemiol. 2019;188(1):231–8.

73. Mokry LE, Ross S, Ahmad OS, Forgetta V, Davey Smith G, Goltzman D, et al. Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLOS Med. 2015;12(8):e1001866–e1001866.

74. Millard LAC, Davies NM, Tilling K, Gaunt TR, Davey Smith G. Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization. PLOS Genet. 2019;15(2):e1007951–e1007951.

75. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42(4):1134–44.

76. Zhao VJ, Luo S, Schooling CM. Sex-specific Mendelian randomization study of genetically predicted insulin and cardiovascular events in the UK Biobank. Commun Biol. 2019;2:332.

77. Burgess S, Seaman S, Lawlor DA, Casas JP, Thompson SG. Missing Data Methods in Mendelian Randomization Studies With Multiple Instruments. Am J Epidemiol. 2011;174(9):1069–76.

78. Liyanage UE, Ong J-S, An J, Gharahkhani P, Law MH, MacGregor S. Mendelian Randomization Study for Genetically Predicted Polyunsaturated Fatty Acids Levels on Overall Cancer Risk and Mortality. Cancer Epidemiol Biomarkers & Prev. 2019;28(6):1015.

79. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.

80. Schmidt AF, Finan C, Gordillo-Marañón M, Asselbergs FW, Freitag DF, Patel RS, et al. Genetic drug target validation using Mendelian randomisation. Nat Commun. 2020;11(1):3255.

81. Labrecque J, Swanson SA. Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. Curr Epidemiol reports. 2018;5(3):214–20.

82. Mountjoy E, Davies NM, Plotnikov D, Davey Smith G, Rodriguez S, Williams CE, et al. Education and myopia: assessing the direction of causality by mendelian randomisation. BMJ. 2018;361:k2022.

83. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. Nat Neurosci. 2018;21(9):1161–70.

84. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l1855.

85. Baldwin J, Pingault J, Schoeler T, Sallis H, Munafo MR. Protecting against researcher bias in secondary data analysis: Challenges and solutions. PsyArXiv. 2020;

86. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. StatMed. 2002;21:1539–1558.

87. Evangelou E, Ioannidis JPA, Patsopoulos NA. Uncertainty in Heterogeneity Estimates in Meta-Analyses. BMJ Br Med J. 2007;335(7626):914–6.

88. Mhatre S, Richmond RC, Chatterjee N, Rajaraman P, Wang Z, Zhang H, et al. The Role of Gallstones in Gallbladder Cancer in India: A Mendelian Randomization Study. Cancer Epidemiol Biomarkers & Prev [Internet]. 2021 Feb 1;30(2):396 LP – 403. Available from: http://cebp.aacrjournals.org/content/30/2/396.abstract

89. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016 Nov;40(7):597–608.

90. Lotta LA, Sharp SJ, Burgess S, Perry JRB, Stewart ID, Willems SM, et al. Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. JAMA. 2016;316(13):1383–91.

91. Lane JM, Jones SE, Dashti HS, Wood AR, Aragam KG, van Hees VT, et al. Biological

and clinical insights from genetics of insomnia symptoms. Nat Genet. 2019;51(3):387–93.

92. Sun Y-Q, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. BMJ. 2019;364:l1042.

93. Wootton RE, Lawn RB, Millard LAC, Davies NM, Taylor AE, Munafò MR, et al. Evaluation of the causal effects between subjective wellbeing and cardiometabolic health: mendelian randomisation study. BMJ. 2018;362:k3788–k3788.

94. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med. 2017;36(11):1783–802.

95. Larsson SC, Bäck M, Rees JMB, Mason AM, Burgess S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: a Mendelian randomization study. Eur Heart J. 2020;41(2):221–6.

96. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89–98.

97. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. Nature. 2018;553(7689):399–401.

98. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866–86.

99. Bowden J, Spiller W, Del Greco M F, Sheehan N, Thompson J, Minelli C, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. Int J Epidemiol. 2018;47(4):1264–78.

100. Hess DR. How to Write an Effective Discussion. Respir Care. 2004;49(10):1238.

101. Docherty M, Smith R. The case for structuring the discussion of scientific papers. BMJ. 1999;318(7193):1224–5.

102. Perneger VT, Hudelson PM. Writing a research article: advice to beginners. Int J Qual Heal Care. 2004;16(3):191–2.

103. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. JAMA. 2017;318(10):947–56.

104. Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017;358:j3542.

105. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol. 2018;33(10):947–52.

106. Davey Smith G. Does schizophrenia influence cannabis use? How to report the influence of disease liability on outcomes in Mendelian randomization studies [Internet]. 2019. Available from: https://targ.blogs.bristol.ac.uk/2019/01/07/does-schizophrenia-influence-cannabis-use-how-to-report-the-influence-of-disease-liability-on-outcomes-in-mendelian-randomization-studies/

107. Campbell DT. Factors relevant to the validity of experiments in social settings. Psychol Bull. 1957;54(4):297–312.

108. Devesa SS, Blot WJ, Fraumeni Jr. JF. Declining Lung Cancer Rates Among Young Men and Women in the United States: A Cohort Analysis. JNCI J Natl Cancer Inst. 1989;81(20):1568–71.

109. Holmes VM, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017;14(10):577–90.

110. Richmond RC, Anderson EL, Dashti HS, Jones SE, Lane JM, Strand LB, et al. Investigating causal relations between sleep traits and risk of breast cancer in women: mendelian randomisation study. BMJ. 2019;365:l2327.

111. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome: systematic review with meta-analysis. Intensive Care Med. 2018;44(10):1603–12.

112. Kesselheim AS, Robertson CT, Myers JA, Rose SL, Gillet V, Ross KM, et al. A Randomized Study of How Physicians Interpret Research Funding Disclosures. N Engl J Med. 2012;367(12):1119–27.

113. Fabbri A, Lai A, Grundy Q, Bero LA. The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review. Am J Public Health. 2018;108(11):e9–16.

114. Henderson GE, Cadigan RJ, Edwards TP, Conlon I, Nelson AG, Evans JP, et al. Characterizing biobank organizations in the U.S.: results from a national survey. Genome Med. 2013;5(1):3.

115. Caulfield T, Burningham S, Joly Y, Master Z, Shabani M, Borry P, et al. A review of the key issues associated with the commercialization of biobanks. J law Biosci. 2014;1(1):94–110.

116. Luo S, Au Yeung SL, Zhao VJ, Burgess S, Schooling CM. Association of genetically predicted testosterone with thromboembolism, heart failure, and myocardial infarction: mendelian randomisation study in UK Biobank. BMJ. 2019;364:l476.

117. Packer M. Data sharing in medical research. BMJ. 2018;360:k510.

118. Bastian H. “They would say that, wouldn’t they?” A reader’s guide to author and sponsor biases in clinical research. J R Soc Med. 2006;99(12):611–4.

119. Journals stating that they follow the ICMJE Recommendations [Internet]. Available from: http://www.icmje.org/journals-following-the-icmje-recommendations/

120. da Costa BRR, Cevallos M, Altman DGDG, Rutjes AWSAWS, Egger M, Costa B Da, et al. Uses and misuses of the STROBE statement: bibliographic study. BMJ Open. 2011;1(1):e000048.

121. Caulley L, Catala-Lopez F, Whelan J, Khoury M, Ferraro J, Cheng W, et al. Reporting guidelines of health research studies are frequently used inappropriately. J Clin Epidemiol. 2020;122:87–94.

122. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Lancet. 2008;371(9619):1149–1150.

123. Allin KH, Nordestgaard BG, Zacho J, Tybjærg-Hansen A, Bojesen SE. C-Reactive Protein and the Risk of Cancer: A Mendelian Randomization Study. JNCI J Natl Cancer Inst. 2010;102(3):202–6.

124. Wade KH, Carslake D, Sattar N, Davey Smith G, Timpson NJ. BMI and Mortality in UK Biobank: Revised Estimates Using Mendelian Randomization. Obesity. 2018;26(11):1796–806.

125. Galuska DA, Will JC, Serdula MK, Ford ES. Are Health Care Professionals Advising Obese Patients to Lose Weight? JAMA. 1999;282(16):1576–8.

Web Appendix with Additional Examples

1. TITLE and ABSTRACT

Indicate Mendelian Randomisation as the study's design in the title and/or the abstract if that is a main purpose of the study.

Example (*One-sample MR*): "Elevated plasma levels of C-reactive protein (CRP), a marker of inflammation, are associated with an increased risk of cancer, but it is unclear whether this association is causal. We examined whether four common single-nucleotide polymorphisms (SNPs) in the CRP gene that are associated with altered plasma CRP levels are causally associated with an increased risk of cancer. The study population included participants in a prospective study (n = 10 215) and a cross-sectional study (n = 36 403) of the adult general population in Denmark, all of whom were genotyped for the CRP SNPs. The association between plasma CRP levels measured by a high-sensitivity turbidimetry assay and the risk of cancer was examined for 8224 participants in the prospective study. The hazard ratio of cancer for a doubling of the plasma CRP level was 1.09 (95% confidence interval [CI] = 1.03 to 1.14). The nine most common genotype combinations of the four CRP SNPs were associated with up to a 72% increase (95% CI = 58% to 87%) in CRP levels but not with an increased risk of cancer. The estimated causal odds ratio for cancer associated with a genetically induced doubling in CRP level was 0.94 (95% CI = 0.81 to 1.08). This finding suggests that elevated CRP levels do not cause cancer." [123]

Example: *Abstract - Two sample MR*

"**OBJECTIVE** To determine whether body mass index, body fat percentage, and waist circumference influence smoking status and intensity.

DESIGN Mendelian randomisation study.

SETTING UK Biobank, with replication of results from the Tobacco and Genetics (TAG) consortium.

PARTICIPANTS European descent participants from the UK Biobank cohort (n=372 791) and the TAG consortium (n=74 035). MAIN OUTCOME MEASURES Risk of current and past smoking, number of cigarettes smoked per day, age of smoking initiation. RESULTS The Mendelian randomisation analysis indicated that each standard deviation increment in body mass index (4.6) increased the risk of being a smoker (odds ratio 1.18 (95% confidence interval 1.13 to 1.23), P<0.001). This association was replicated in the TAG consortium data (1.19 (1.06 to 1.33), P=0.003). Furthermore, each standard deviation increment in body mass index was estimated to increase smoking intensity by 0.88 cigarettes per day (95% confidence interval 0.50 to 1.26, P<0.001) in UK Biobank and 1.27 cigarettes per day in the TAG consortium (0.46 to 2.07, P=0.002). Similar results were also seen for body fat percentage and waist circumference in both UK Biobank and the TAG consortium data.

CONCLUSIONS These results strongly suggest that higher adiposity influences smoking behaviour and could have implications for the implementation of public health interventions aiming to reduce the prevalence of these important risk factors." [46]

Example: ***Abstract - Embedded MR (as part of a larger analysis)***

"**OBJECTIVE:** To identify the genetic determinants of fracture risk and assess the role of 15 clinical risk factors on osteoporotic fracture risk. DeSiGN Meta-analysis of genome wide association studies (GWAS) and a two-sample mendelian randomisation approach.

SETTING: 25 cohorts from Europe, United States, east Asia, and Australia with genome wide genotyping and fracture data.

PARTICIPANTS: A discovery set of 37 857 fracture cases and 227 116 controls; with replication in up to 147 200 fracture cases and 150 085 controls. Fracture cases were defined as individuals (>18 years old) who had fractures at any skeletal site confirmed by medical, radiological, or questionnaire reports. Instrumental variable analyses were performed to estimate effects of 15 selected clinical risk factors for fracture in a two- sample mendelian randomisation framework, using the largest previously published GWAS meta-analysis of each risk factor.

RESULTS: Of 15 fracture associated loci identified, all were also associated with bone mineral density and mapped to genes clustering in pathways known to be critical to bone biology (eg, SOST, WNT16, and ESR1) or novel pathways (FAM210A, GRB10, and ETS2). Mendelian randomisation analyses showed a clear effect of bone mineral density on fracture risk. One standard deviation decrease in genetically determined bone mineral density of the femoral neck was associated with a 55% increase in fracture risk (odds ratio 1.55 (95% confidence interval 1.48 to 1.63; P=1.5×10−68). Hand grip strength was inversely associated with fracture risk, but this result was not significant after multiple testing correction. The remaining clinical risk factors (including vitamin D levels) showed no evidence for an effect on fracture.

CONCLUSIONS: This large scale GWAS meta-analysis for fracture identified 15 genetic determinants of fracture, all of which also influenced bone mineral density. Among the clinical risk factors for fracture assessed, only bone mineral density showed a major causal effect on fracture. Genetic predisposition to lower levels of vitamin D and estimated calcium intake from dairy sources." [38]

4.d) For each exposure, outcome and other relevant variables, describe methods of assessment and diagnostic criteria for diseases.

Example: ***Categorical exposure or outcome***

"The outcome of the study was prevalent type 2 diabetes, defined consistent with validated algorithms developed for UK Biobank. (Eastwood et al Plos One 2016) Participants were classified as cases if they met the following 2 criteria: (1) self-reported type 2 diabetes diagnosis or self-reported diabetes medication at nurse interview or at digital questionnaire, or electronic health record consistent with type 2 diabetes (International Statistical Classification of Diseases and Related Health Problems Tenth Revision code E11); and (2) age at diagnosis older than 36 years or use of oral antidiabetic medications (to exclude likely type 1 diabetes cases). Controls were participants who (1) did not self-report a diagnosis of diabetes of any type, (2) did not take any diabetes medications, and (3) did not have an electronic health record of diabetes of any type." [43]

Example: ***All-cause and cause-specific mortality***

"Data from death certificates were sent to UK Biobank on a quarterly basis provided by the National Health Service (NHS) Information Centre for participants from England and Wales and by NHS Central Register, Scotland for participants from Scotland. More detailed information on mortality are available at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=115559. The death certificates include the

disease or condition stated to be the underlying cause of death, as well as other conditions, diseases, injuries or events contributing to death but not related to the disease or condition causing it. Data were provided as date of death (DoD), an integer value for age of death (AoD) and underlying (primary) cause of death in International Classification of Diseases (ICD)-10 codes for all deaths that occurred between the 10/05/2006 and 16/02/2016. Rather than using the integer value of AoD from the death certificate, a more precise measure of AoD was derived by adding the time interval between date of initial assessment and DoD (in days) to the participant's age at initial assessment. All participants who were not recorded as dead by the 16th of February 2016 were assumed to still be alive. The ICD-10 codes were categorised into all-cause and cause-specific mortality as presented in Table S1a. As of August 2017 (date of extraction for all data), there were 14,417 total deaths in the entire UK Biobank dataset that had occurred up to 16th of February 2016 (Table S1a for the whole sample and Table S1b for males and females), which remains the most updated data on mortality.

For the purposes of this study, the primary outcomes of focus were as follows: all-cause mortality and mortality from all cardiovascular diseases and those specifically due to coronary heart disease, stroke, aortic aneurysm and any other cardiovascular diseases; overall cancer and those specifically due to cancers of the lung, colorectum, prostate (men only), breast cancer (women only, separated into pre- and post-menopausal occurrences), pancreas, ovaries (women only), endometrium (women only), stomach, oesophagus, skin (malignant melanoma), kidney, bladder, brain, lymphatic system and all other cancers; and external causes." [124]

8. Sensitivity analyses and additional analyses

Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations).

Example: "Although the assumption cannot be verified, the sensitivity or robustness of estimates to violations of this assumption can be evaluated using traditional bias analysis techniques. A particularly simple, easy-to-use approach to evaluate the sensitivity of estimates to confounding is the E-value. […] The E-value [is] reported for both the estimate and the limit of the confidence interval closest to the null." [64]

10. Descriptive data

a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider the use of a flow diagram.

Example: "UK Biobank is a prospective cohort that recruited more than 500 000 men and women aged 40- 96 years between 2006 and 2010, and collected anthropometric, health, and lifestyle data, as well as biological samples. Of 487 409 individuals who were genotyped in UK Biobank, we used data for 372 791 European descent participants with valid adiposity and smoking behaviour measures at recruitment. European background was genetically assessed through principal component analyses of data from genome wide association studies. Sample quality control steps are given in the supplementary methods." [46]

Example: "Supplementary figure 1 shows the exclusion criteria for the main UK Biobank analyses, and supplementary figure 2 shows the exclusion criteria for the genome-wide association studies carried out for systolic blood pressure and smoking. White British participants were defined by using both self-reported questionnaire data and similar genetic ancestry to the European ancestry principal components computed from the 1000 genomes project." [84]

Supplementary Figure 2: Flow chart for exclusions made in UK Biobank for use in SBP and smoking GWAS analyses

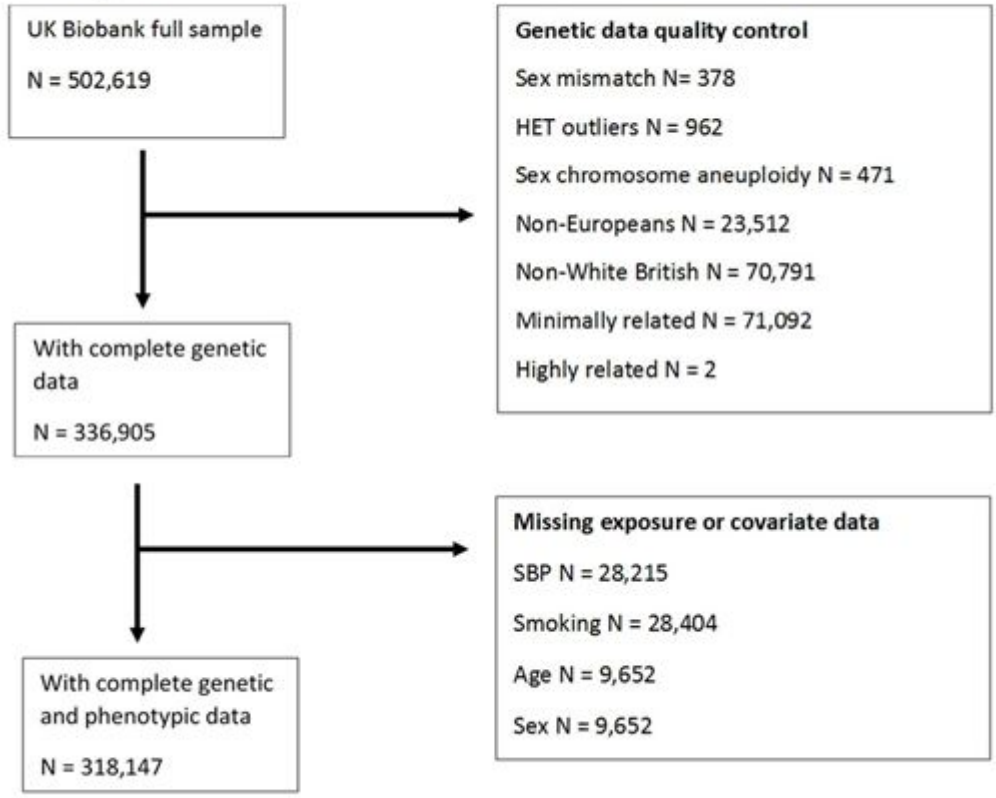


Figure reproduced with permission from Carter et al., 2019.[84]

15. Limitations

Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and your efforts to address them.

Example: "Since the prevalence of counseling increases with increasing levels of obesity, our estimates may overestimate the true prevalence. Telephone surveys also may overestimate the true prevalence of counseling. Although persons without telephones have similar levels of overweight as persons with telephones, persons without telephones tend to be less educated, a factor associated with lower levels of counseling in our study. Also, of concern is the potential bias caused by those who refused to participate as well as those who refused to respond to questions about weight. Furthermore, because the data were collected cross-sectionally, we cannot infer that counseling preceded a patient's attempt to lose weight." [125]