

Methods

Tools for assessing quality and risk of bias in Mendelian randomization studies: a systematic review

Francesca Spiga ^{1,2*} Mark Gibson ^{2,3} Sarah Dawson,¹
Kate Tilling,^{1,2} George Davey Smith,^{1,2} Marcus R Munafò^{2,3} and
Julian PT Higgins ^{1,2}

¹Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, ²Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK and ³School of Psychological Science, University of Bristol, Bristol, UK

*Corresponding author. Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK. E-mail: f.spiga@bristol.ac.uk

Received 8 February 2022; Editorial decision 10 June 2022; Accepted 29 June 2022

Abstract

Background: The use of Mendelian randomization (MR) in epidemiology has increased considerably in recent years, with a subsequent increase in systematic reviews of MR studies. We conducted a systematic review of tools designed for assessing risk of bias and/or quality of evidence in MR studies and a review of systematic reviews of MR studies.

Methods: We systematically searched MEDLINE, Embase, the Web of Science, preprints servers and Google Scholar for articles containing tools for assessing, conducting and/or reporting MR studies. We also searched for systematic reviews and protocols of systematic reviews of MR studies. From eligible articles we collected data on tool characteristics and content, as well as details of narrative description of bias assessment.

Results: Our searches retrieved 2464 records to screen, from which 14 tools, 35 systematic reviews and 38 protocols were included in our review. Seven tools were designed for assessing risk of bias/quality of evidence in MR studies and evaluation of their content revealed that all seven tools addressed the three core assumptions of instrumental variable analysis, violation of which can potentially introduce bias in MR analysis estimates.

Conclusion: We present an overview of tools and methods to assess risk of bias/quality of evidence in MR analysis. Issues commonly addressed relate to the three standard assumptions of instrumental variables analyses, the choice of genetic instrument(s) and features of the population(s) from which the data are collected (particularly in two-sample MR), in addition to more traditional non-MR-specific epidemiological biases. The identified tools should be tested and validated for general use before recommendations can be made on their widespread use. Our findings should raise awareness about the

importance of bias related to MR analysis and provide information that is useful for assessment of MR studies in the context of systematic reviews.

Key words: Mendelian randomization, genetic instrument, bias, tool, guideline, risk-of-bias assessment

Key Messages

- Mendelian randomization (MR) analyses are potentially a powerful approach to making causal inferences in observational studies, but they are built on important assumptions and are not immune to bias.
- We identified 14 tools for the conduct, evaluation and/or reporting of MR studies. We focused on the evaluation tools and most of these contained items addressing validity of the three core instrumental variables assumptions.
- We also examined systematic reviews of MR studies and protocols for systematic reviews. Only a small proportion of the reviews conducted a risk-of-bias assessment although most included a narrative description of MR-related bias and limitations. Most of the protocols planned to conduct an assessment, although fewer than half of them planned methods specifically for MR studies.
- The tools we identified should provide a useful source of information on what bias/limitations reviewers should be aware of when reading or systematically reviewing MR studies.

Introduction

Mendelian randomization (MR) is an analytic approach used to make causal inference in observational studies.¹ In MR analysis, genetic variants are generally used as instrumental variables (genetic instruments) to estimate the causal effect of a modifiable trait (the causal factor or ‘exposure’) on another trait (the factor or condition that the exposure is hypothesized to influence or ‘outcome’).² Causal inference using MR analysis is based on the notion that genetic variants are randomly inherited from parents to offspring in a way that is comparable to participants being randomly allocated to each experimental group in a randomized-controlled trial (RCT).³ In a within-sibship analysis randomization is almost exact⁴ and MR was introduced through this hypothetical approach,¹ but until recently large-scale data were not available to conduct such analyses and the approximate randomization in population-level data (adjusted for potential population stratification) has been the main approach.³ Thus, the key advantage of using a MR approach is the potential to reduce bias due to residual confounding and reverse causation, which are often limitations in other types of observational studies.^{1,5} MR was introduced as a way of strengthening causal inference regarding the effects of modifiable exposures studied in conventional observational epidemiological studies.^{1,6}

As for instrumental variables analyses in general, the validity of an estimate from a MR analysis relies on the

genetic instrument satisfying three core assumptions:¹ the genetic instrument must be associated with the exposure (IV1-relevance),² there are no unmeasured confounders of the genetic instrument–outcome association (IV2-independence) and³ the genetic instrument–outcome association must be mediated entirely via the exposure (IV3-exclusion restriction). Additional assumptions, which are variations of a fourth IV assumption (IV4),^{2,7} may be required for some inferences. Versions of these include (i) the association of the genetic instrument and the exposure and the effect of the exposure on the outcome are the same for all participants in the sample (homogeneity); (ii) the genetic instrument does not modify the effect of the exposure on the outcome within levels of the exposure and for all levels of the exposure (no effect modification); (iii) the direction of the effect of the exposure on the outcome is the same for all participants in the sample (monotonicity).⁸ Finally, to consider that the findings inform intervention strategies it must be assumed that the differences in an exposure induced by the genetic instrument will produce the same downstream effects on health outcomes as differences in the exposure produced by environmental influences (gene–environment equivalence assumption).^{2,9} The validity of two-sample MR studies, in which different samples are used to estimate the genetic instrument–exposure and genetic instrument–outcome associations, relies on additional assumptions that the samples are independent (i.e. do not overlap): the samples are from the same underlying

population (e.g. same age range, same ancestry)¹⁰ (thus with the genetic variants being equally distributed in the two populations)¹¹ and the genetic variants are harmonized (i.e. they are in the same direction in the two samples).¹⁰

Some of the specific biases that have been articulated in relation to MR studies include biases emerging from the genetic instrument (e.g. weak instrument bias,¹² bias due to horizontal pleiotropy,¹³ bias due to linkage disequilibrium,⁵ bias due to developmental compensation)¹ and biases related to the population from which the data are collected (e.g. bias due to population stratification, assortative mating, dynastic effect and parent of origin effect,^{1,14,15} bias due to sample overlap in two-sample MR).¹⁶ Using weak instruments in MR analysis creates a problem in relation to IV1 and can lead to estimates biased towards the confounded exposure–outcome association (in one-sample MR) or towards the null (in two-sample MR). Failure to adjust for population structure and familial effects can introduce confounding in a way that is similar to lack of randomization in a RCT and relates to IV2.¹⁴ Horizontal pleiotropy leads to violation of IV3. Some problems can lead to violation of more than one IV assumption; e.g. linkage disequilibrium can introduce both horizontal pleiotropy (IV3)¹⁷ and confounding (IV2).⁵ In addition, biased estimates can arise from other more general types of bias, including measurement/classification biases, selection biases (including those due to missing data and to collider bias) and reporting biases.

Since the initial detailed exposition of MR in epidemiology in 2003,¹ its use has increased very considerably and with this has come a parallel increase in systematic reviews of MR studies. One important component of a systematic review (and meta-analysis) is the evaluation of the quality of evidence reported in each study included. This is increasingly achieved by assessing risk of bias through a structured framework. Although numerous tools for risk-of-bias assessment in studies of interventions have been developed for both RCTs¹⁸ and non-randomized studies of intervention,¹⁹ and are widely used, there is no widely agreed tool for assessing MR studies.

In this systematic review we sought to identify and examine structured frameworks used to assess risk of bias (or quality more generally) in MR studies. Specifically, we undertook a comprehensive and objective review of tools for the systematic evaluation of MR studies; identified and summarized tools for assessing the conduct and/or reporting of MR studies to examine what bias-related features they covered; and undertook an examination of how risk of bias in MR studies has been assessed in systematic reviews to date.

Methods

Eligibility criteria

For the review of existing tools, we sought structured guidelines, checklists and other tools aimed at comprehensive evaluation of the conduct, evaluation and/or reporting of MR studies or structured guidance through the steps of conducting or reporting an MR study. For the review of systematic reviews, we examined articles describing systematic approaches to collating and summarizing MR studies within a field or more generally. We considered a systematic review any article in which the authors (i) undertook a bibliographic database search (e.g. in MEDLINE and/or other databases); and (ii) provided a table describing each of the included studies. We included full reports (e.g. full-text articles) and protocols, but not conference abstracts (unless an associated full-text report could be identified). We regarded any article in which genetic variants have been described or used as instrumental variables as relevant to our review.

Searches

We performed systematic electronic searches in (i) MEDLINE (Ovid), Embase (Ovid) and the Web of Science (from inception to 30 June 2021) for published peer-reviewed articles and (ii) bioRxiv and medRxiv for preprint articles (last search 1 July 2021). We implemented specific searches to identify articles describing tools (Search 1), systematic reviews (Search 2) and protocols for systematic reviews (Search 3). To identify systematic reviews, we also searched Epistemonikos and for information on ongoing reviews we searched PROSPERO and Open Science Framework (OSF) Registries (last search 1 July 2021). To identify additional articles and protocols (missed from the bibliographic database searches), we searched Google Scholar, examined references of included studies and performed forward citation searches (Google Scholar) to identify articles citing included studies. Details of search strategies are reported in [Supplementary data](#) (available as [Supplementary data](#) at *IJE* online).

Study selection

Search results were managed using EndNote 20 and Excel. Titles and abstracts were screened by one review author (F.S.) using Rayyan app (www.rayyan.ai). The full text of selected studies was retrieved and assessed for eligibility and inclusion in the review. Full-text screening was performed independently by two review authors (F.S. and M.G.) and disagreements between the two reviewers were

resolved through discussion. Any structured tool identified from the review of systematic reviews was incorporated into the review of tools.

Data extraction

An extraction form was used to extract the data from the articles selected for inclusion. For each sub-review, a pilot data extraction was performed and a finalized data extraction form was compiled. From each article, the following general information was extracted by one review author (F.S.): first author(s) name and year of publication, type of report (full-text article or conference abstract), type of article (e.g. tool, systematic review, protocol of systematic review) and complete reference. In addition, information specific to the sub-reviews was extracted as follows:

Review of tools: number of tools within the article, purpose of the tool (i.e. conducting, evaluating or reporting), structure of the tool (e.g. guide, dictionary, checklist) and for the evaluating tools only, specific objectives of the article, other tools used as template, number of domains and items (or questions) and specific content of each item within each tool. We extracted information only about tools designed specifically for MR studies.

Review of systematic reviews: review topic, whether only MR studies were included, number of included MR and non-MR studies, whether a systematic assessment of risk of bias was undertaken (or proposed if a protocol) and, if applicable, whether a structured tool was used, what biases were addressed, how biases were addressed, if a narrative description of MR-specific bias was reported and what biases were narratively addressed. We also evaluated whether a systematic assessment of the quality of evidence supporting a causal effect reported by individual MR studies was undertaken and, if applicable, what approaches were used.

Data analysis and reporting

We report our findings using structured summary tables and narrative descriptions. For the tools identified in the first sub-review that were aimed at the evaluation of an MR study, we tabulate the items addressed by the different tools. Where an item contained multiple questions, we separate these and tabulate each question as a single item. We mapped items across tools to examine how similar biases were addressed by different evaluating tools and to convey how many of the tools addressed each bias. Specifically, we classified each item into a broad bias/topic domain and then we assigned each item to a specific bias/topic within that domain and determined the numbers of items allocated to each bias domain and to specific MR bias/

topic. For the systematic reviews, we tabulate the methods of risk-of-bias and/or quality-of-evidence assessment in MR studies and the MR-relevant bias addressed either by the method of assessment used or within a narrative description. For protocols of systematic reviews, we tabulate the proposed methods of assessment of risk of bias/quality of evidence in MR studies. Data extraction, narrative synthesis and tabulations were performed by one reviewer (F.S.).

Results

Tools for the conduct, evaluation and reporting of MR studies

In total, 363 records were identified from the searches (352 from database searches and 11 from other searches) of which 19 were retrieved for full-text screening. The inclusion criteria were met by 13 articles (reporting 14 tools) that are included in this review. A flow diagram of the identification, screening and inclusion of articles is shown in [Figure 1](#). Of the 13 included articles, 6 were identified from searches of electronic databases of peer-reviewed articles and 4 from searches of preprints archives and Google Scholar, 2 from cited references, 2 from searches of systematic reviews (Search 2) and 1 from searches of protocols of systematic reviews (Search 3). A list of all the included tools is reported in [Supplementary Table S1](#) (available as [Supplementary data](#) at *IJE* online) and the six studies that did not meet the criteria for inclusion are listed in [Supplementary data](#) (available as [Supplementary data](#) at *IJE* online).

Of the 14 tools included, 8 tools were designed for single use in a specific systematic review (7 reviews and 1 protocol) and 6 tools were proposed for future use for the conduct, evaluation and/or reporting of MR studies in general or within the context of a systematic review. Of the 14 identified tools, 8 tools had a single purpose, of which 4 were aimed at the conduct of MR studies, 3 were aimed at the reporting of MR studies and 1 was aimed at the evaluation of MR studies. The remaining six tools had two purposes: evaluation and reporting of MR studies.

Details of the seven tools designed (or used) for evaluation of MR studies are reported in [Table 1](#). Of these, Burgess,²⁰ Davies,²¹ Grau-Perez²² and Treur³¹ were structured by domains and items, whereas Kuźma,²⁴ LS Lee²⁶ and Mamluk²⁸ were structured by items only. The number of domains within the first four evaluating tools ranged from 5 to 9, with a median of 6 and a total of 26 domains across the tools. The number of items in the evaluating tools ranged from 5 to 28, with a median of 19 and a total of 121 items across all the tools.

We conducted a thorough analysis of the structure and content of the evaluating tools by classifying each item into

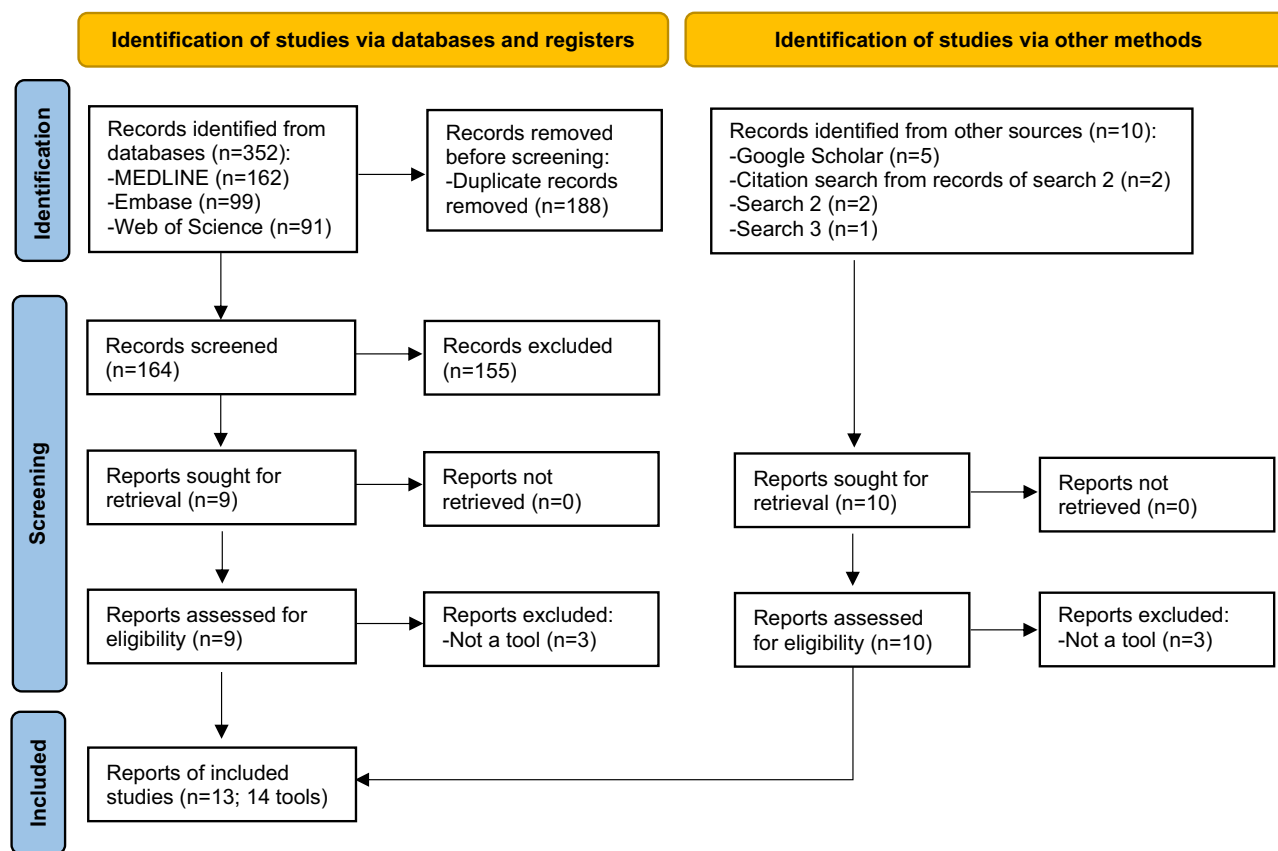


Figure 1 Flow diagram of identification, screening and inclusion of articles containing tools for the conduct, evaluation and reporting of Mendelian randomization studies

a bias/topic domain and then we assigned each item to a specific bias/topic. We found that of the 121 items among all evaluating tools, 81 items were designed to evaluate risk of bias in MR studies and 44 items were designed to address other aspects of the MR analysis (4 items were designed to address both evaluation of risk of bias and other aspects of MR analysis). Of the 81 items designed to evaluate MR studies, 77 addressed only one bias and 4 addressed multiple biases.

Details of the biases addressed by each evaluating tool are reported in [Table 2](#). Of the 81 items addressing bias, 32 related to the three core IV assumptions. Ten items in 7 tools addressed bias related to the relevance assumption (IV1), 8 items in 6 tools addressed bias related to the independence assumption (IV2) and 14 items in 7 tools addressed bias related to the exclusion restriction assumption (IV3). In addition, 11 items in 4 tools addressed bias related to the selection of the genetic instrument and 14 items in 6 tools addressed bias related to the selection of the population(s) or sample(s). Five items in 4 tools addressed bias related to sensitivity analysis, 19 items in 3 tools addressed bias related to measurement errors and misclassification, 2 items in 1 tool addressed bias due to missing data, 4 items in 3 tools addressed bias due to other

types of confounding and 2 items in 1 tool addressed other sources of bias. We provide details of the 44 items addressing other aspects of the MR analysis, including items addressing the reporting of MR analysis, in [Supplementary Table S2](#) (available as [Supplementary data](#) at *IJE* online). Among these, we found that two items in one tool addressed clinical implications of the MR results; three items in three tools addressed the choice of data set(s); four items in three tools addressed the genetic instrument; six items in two tools addressed the interpretation of the MR analysis results; five items in three tools addressed the MR rationale; six items in three tools addressed the MR results; four items in three tools addressed precision of the results; two items in one tool addressed the selection of the population(s) or sample(s); and seven items in four tools addressed the statistical analysis.

In addition to the evaluating tools, we identified three tools aimed at reporting and four tools aimed at conducting MR studies. All seven tools contained items addressing bias in MR analysis and details of the content of the items is reported in [Supplementary Table S3](#) (available as [Supplementary data](#) at *IJE* online). The number of domains ranged from 3 to 6 in the reporting tools and from 5 to 10 in the conducting tools; the number of items

Table 1 Details of tools designed/used for assessing risk of bias/evaluating quality in Mendelian randomization studies

Author	Objectives of the article	Tool used as template/reference to other tools or articles relevant to MR	N of domains	N of items or questions
Burgess 2020 ²⁰	To provide guidelines for performing MR investigations. To provide advice on which analyses to perform in a MR investigation		9	22
Davies 2018 ²¹	To provide explanations of core concepts and recent developments in MR methods		6	19
Grau-Perez 2019 ²²	To conduct a systematic review of MR studies evaluating the causal role of environmentally responsive DNAm changes on the development of health states	Boef <i>et al.</i> ²³	6	28
Kużma 2018 ²⁴	To conduct a systematic review of MR studies investigating causal relationships between risk factors and global cognitive function or dementia	Q-Genie ²⁵	–	11
Lee 2020 ²⁶	To perform an updated systematic review and meta-analysis of MR that will provide further insight into the causative factors of dementia	Davies <i>et al.</i> ²¹ Grover <i>et al.</i> ²⁷ Burgess <i>et al.</i> ²⁰	–	11
Mamluk 2020 ²⁸	To conduct a systematic review of human studies that used experimental data or alternative analytical methods to determine the causal effects of maternal alcohol consumption in pregnancy on offspring outcomes at birth and later in life	Glymour <i>et al.</i> ²⁹ Lawlor <i>et al.</i> ¹⁵ Taylor <i>et al.</i> ³⁰	–	5
Treur 2021 ³¹	To review evidence from studies that applied MR to assess causal effects between poor mental health and substance use	STROBE-MR ^{32,33}	5	25

DNAm, DNA methylation; MR, Mendelian randomization.

ranged from 7 to 61 in the reporting tools and from 18 to 26 in the conducting tools. Among the reporting tools, all three tools contained items addressing the three IV core assumptions; Boef²³ contained items addressing linkage disequilibrium and canalization; Davey Smith³² contained items addressing homogeneity and sample overlap (in two-sample MR); Loi³⁴ contained items addressing linkage disequilibrium and heteroscedasticity. Among the conducting tools, Burgess,²⁰ Grover²⁷ and Lawlor³⁵ contained items addressing the three IV core assumptions and variant harmonization; in addition, Burgess²⁰ contained one item addressing the homogeneity assumptions and Grover²⁷ and Lawlor³⁵ contained items addressing sample overlap; Swerdlow³⁶ contained items addressing linkage disequilibrium and horizontal pleiotropy.

Systematic reviews of MR studies

Completed reviews

A total of 2036 records were identified from Search 2 (for systematic reviews) (2025 from database searches and 11 from other searches) of which 143 were retrieved for full-text screening and the inclusion criteria were met by 38 articles

(35 full-text articles and 3 conference abstracts linked to included articles) reporting 35 reviews that are included in this synthesis. A flow diagram of identification, screening and inclusion of studies is shown in [Figure 2](#). A list of included reviews is reported in [Table 3](#) and the 104 studies that did not meet the criteria for inclusion are listed in [Supplementary data](#) (available as [Supplementary data](#) at *IJE* online).

Of the 35 included reviews, 25 were systematic reviews and 10 were umbrella reviews. Of the 35 included reviews, 29 addressed a clinical question (i.e. included studies on the causal effect of an exposure vs an outcome) and 6 reviews addressed a methodological question (e.g. the status of reporting in MR studies); 17 reviews reported MR studies only and the other 18 reported both MR and non-MR studies; the number of MR studies ranged between 1 and 231 with a median of 18 studies. Of the 35 included reviews, 14 conducted an assessment of either risk of bias or quality of the evidence: 6 reviews conducted risk-of-bias assessments only, 5 reviews conducted quality-of-evidence assessments only and 3 did both. Details of the risk-of-bias and quality-of-evidence assessment in individual MR studies used in these 14 reviews are reported in [Supplementary Table S4](#) (available as [Supplementary data](#) at *IJE* online).

Table 2 Details of specific Mendelian randomization bias and limitation addressed by items or questions within each assessing tool

Bias (or topic) domain	Specific bias or topic addressed by tool	Burgess 2020	Davies 2018	Grau-Perez 2019	Kuźma 2018	Lee 2020	Mamluk 2020	Treur 2021	Total items
IV1-Relevance	Choice of variants	Yes							1
	Weak instrument bias		Yes	Yes	Yes	Yes	Yes	Yes	9
IV2-Independence	Choice of variants	Yes							1
	Confounding		Yes	Yes	Yes	Yes ^a	Yes ^b	Yes	7
	Population stratification						Yes		1
IV3-Exclusion restriction	Choice of variants	Yes							1
	Horizontal pleiotropy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	13
Genetic instrument	Choice of variants	Yes							3
	Construction of genetic score							Yes	1
	Variants harmonization	Yes	Yes			Yes		Yes	4
	Linkage disequilibrium	Yes	Yes			Yes			3
Population/sample	Samples overlap ^a	Yes	Yes			Yes		Yes	5
	Population heterogeneity ^c	Yes	Yes	Yes		Yes	Yes	Yes	6
	Choice of controls			Yes					1
	Selection bias			Yes					1
Sensitivity analysis	Evidence of robustness	Yes	Yes			Yes		Yes	5
Measurement error	Exposure measurement error/ misclassification			Yes	Yes			Yes	13
	Outcome measurement error/ misclassification			Yes	Yes			Yes	6
Missing data	Missing data			Yes					2
Other confounding	Non-MR-specific confounding			Yes ^d	Yes		Yes ^e		4
Other sources of bias	Traditional epidemiologic biases (i.e. non-MR-specific)				Yes ^f				2

^aConfounding of the genetic instrument–outcome association.

^bConfounding of the genetic instrument–exposure association and of the genetic instrument–outcome association.

^cIn two-sample MR analysis.

^dConfounding factors for the exposure–outcome or mediator–outcome association.

^eIncluded confounders in the IV analysis.

^fSurvival and diagnostic bias. IV1, instrumental variable assumption 1; IV2, instrumental variable assumption 2; IV3, instrumental variable assumption 3; MR, Mendelian randomization.

A structured risk-of-bias tool for was used in five reviews: four of these (Grau-Perez,²² Kuźma,²⁴ Mamluk²⁸ and Treur³¹) used tools developed specifically for risk-of-bias assessment in MR studies that are included in the above sub-review of tools (see [Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online, and [Table 2](#)); the fifth, Cheng,⁴⁵ used the Newcastle-Ottawa Scale (NOS) for cohort studies,⁶⁹ which was not specifically developed for MR studies. Four further reviews conducted risk-of-bias assessments but did not use a structured tool: Markozannes⁵⁷ and X Zhang⁶⁷ assessed horizontal pleiotropy; Pearson-Stuttard⁵⁹ addressed the selection of the genetic instrument(s); and Riaz^{61,62} conducted evaluation of the three core assumptions.

Of the eight reviews that conducted a quality-of-evidence assessment, Markozannes⁵⁷ and Pearson-Stuttard⁵⁹ used a structured method based on statistical significance of the effect estimate and X Zhang⁶⁷ used a structured method

based on a combination of statistical significance of the effect estimate, statistical power and evidence of bias due to directional pleiotropy. Among the other five reviews in which a structured method was not used, Bochud⁴³ based the assessment of quality of evidence on the strength of the genetic variant; Firth⁴⁷ based the assessment on the results of the statistical analysis, the use of sensitivity analysis and test for bidirectional effects; Kim⁵¹ based the assessment on statistical power; Kohler⁵² based the assessment on the proportion of variance in risk factors explained by genetic instruments used; and Li⁵⁵ based the assessment on the statistical significance of the effect estimate and the statistical power.

Of the 35 reviews included, 28 reported a general narrative description of potential bias and limitation in MR studies. Details of specific biases addressed narratively within these systematic reviews are reported in [Supplementary Table S3](#) (available as [Supplementary data](#) at *IJE* online).

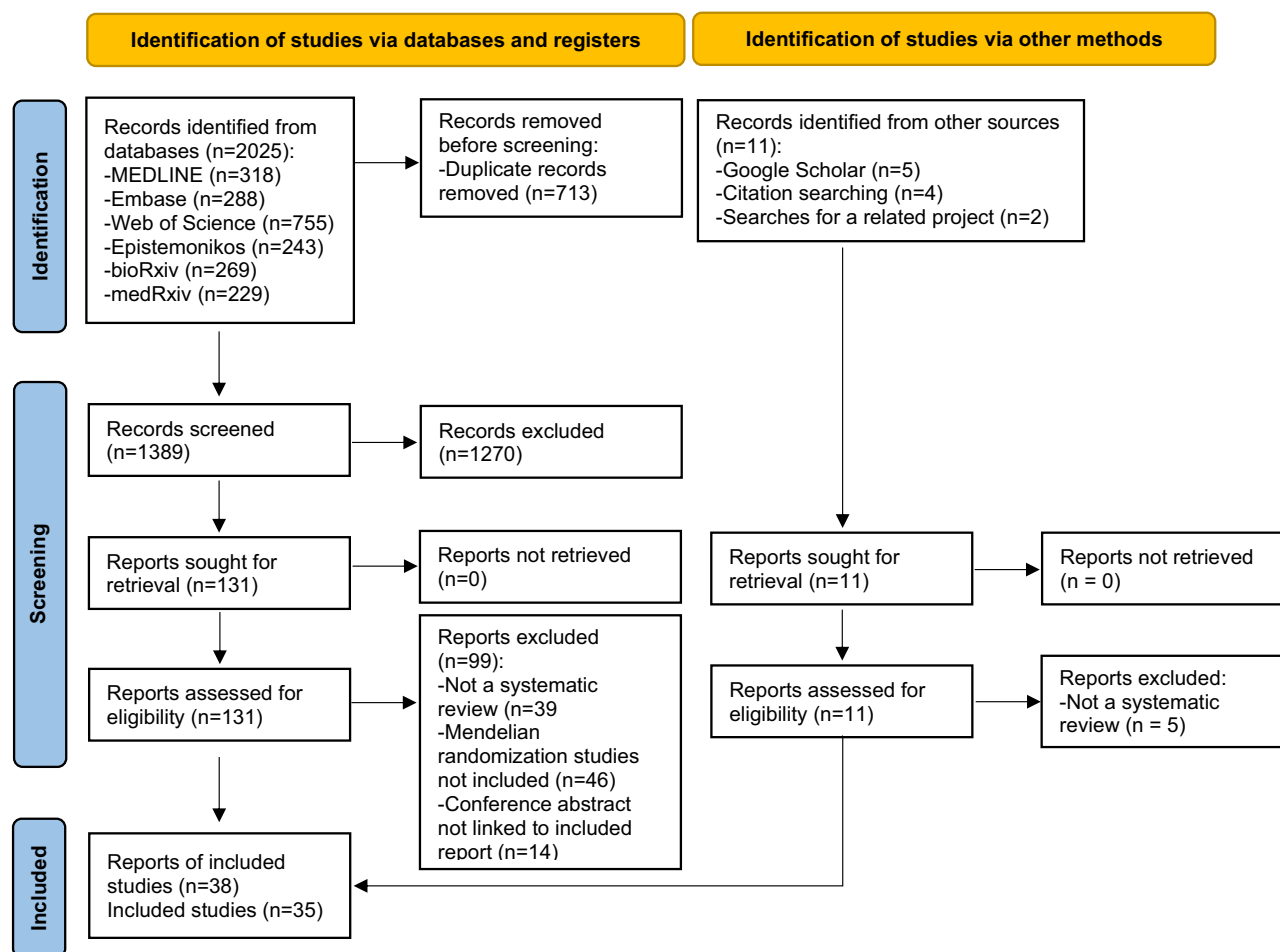


Figure 2 Flow diagram of identification, screening and inclusion of articles containing systematic reviews (and meta-analysis) of Mendelian randomization studies

Of these 28 reviews, 20 addressed bias related to the IV1 assumption (i.e. weak instrument bias), 16 reviews addressed bias related to the IV2 assumption (i.e. confounding, population stratification, assortative mating, dynastic effect and parent of origin effect)¹⁴ and 24 reviews addressed bias related to the IV3 assumption (i.e. horizontal pleiotropy). In addition, 17 reviews addressed bias related to the selection of the genetic instrument (i.e. linkage disequilibrium, Winner's course bias, segregation distortion, monotonicity and homogeneity), 6 reviews addressed bias related to the selection of the population or sample (i.e. population heterogeneity and selection bias), 8 reviews addressed bias due to canalization and 4 reviews addressed bias due to measurement errors or misclassification. In addition to bias, we also evaluated whether other MR-relevant topics were narratively described and we found that 11 reviews addressed precision of the results (i.e. low statistical power or sample size), 5 reviews addressed reverse causation (or bidirectionality), 3 reviews addressed the inability to assess non-linear

associations, 2 reviews addressed statistical analysis and lack of genetic instrument, respectively, and 1 review addressed inability to assess dose-response estimations.

Protocols for systematic reviews

Our final search for protocols of systematic reviews (Search 3) identified 65 protocols (57 from database searches and 8 from other searches, including 1 from Search 2) of which 15 were excluded because inclusion of MR studies was not specified or MR studies were specified in the exclusion criteria. A flow diagram of identification, screening and inclusion of protocols of systematic reviews is shown in Figure 3. Two protocols for the same review were identified from different sources for five reviews so a total of 45 study protocols were included in this part of the review. A list of included protocols with details of the method used by each of study is reported in Table 4 and the 15 protocols that did not meet the criteria for inclusion are listed in Supplementary data (available as

Table 3 List of included systematic reviews reporting one or more Mendelian randomization studies

Study ID	Type of article	Topic of the review	Were only MR studies included?	N of MR studies/N of non-MR studies	Risk-of-bias assessment in individual MR studies? If Yes, was a structured tool used?	Name and/or description of risk-of-bias assessment method	Evidence of causal effect assessment in individual MR studies? If Yes, was a structured method used?	Description of evidence of causal effect assessment method	Narrative description of MR-specific bias
Abbasi 2015 ³⁷	Systematic review and MR analysis	MR studies of biomarkers and T2D	Yes	28/0	No	N/A	No	N/A	Yes
Abbasi 2016 ³⁸	Systematic review	Biomarkers and T2D	No	17/122	No	N/A	No	N/A	Yes
Belbasis 2020 ³⁹	Umbrella review	Risk factors and peripheral biomarkers for schizophrenia and other psychotic disorders	No	5/36	No	N/A	No	N/A	No
Belbasis 2018 ⁴⁰	Umbrella review	Risk factors of multiple sclerosis	No	6/9	No	N/A	No	N/A	Yes
Bellou 2018 ⁴¹	Umbrella review	Environmental risk factors and biomarkers for T2D	No	22/86	No	N/A	No	N/A	Yes
Bergmans 2021 ⁴²	Systematic review	Comorbid depression and T2D	No	4/12	No	N/A	No	N/A	Yes
Bochud 2010 ⁴³	Literature review on MR methods, applications and limitations	MR studies	Yes	38/0	No	N/A	Yes/No	Strength of genetic variant	Yes
Boef 2015 ^{a,23}	Systematic review	Methodology used in MR analysis	Yes	179/0	No	N/A	No	N/A	Yes
Carnegie 2020 ⁴⁴	Literature review on MR methods, applications and limitations and systematic review of MR studies	MR in Nutritional psychiatry	Yes	26/0	No	N/A	No	N/A	Yes
Cheng 2020 ⁴⁵	Systematic review and meta-analysis	Puberty timing and T2D and/or impaired glucose tolerance	No	1/27	Yes, Yes	Newcastle-Ottawa Scale ⁶⁹	No	N/A	No
Diemer 2021 ⁴⁶	Systematic review	Prenatal environment and offspring outcomes	Yes	43/0	No	N/A	No	N/A	Yes

(Continued)

Table 3 Continued

Study ID	Type of article	Topic of the review	Were only MR studies included?	N of MR studies/N of non-MR studies	Risk-of-bias assessment in individual MR studies? If Yes, was a structured tool used?	Name and/or description of risk-of-bias assessment method	Evidence of causal effect assessment in individual MR studies? If Yes, was a structured method used?	Description of evidence of causal effect assessment method	Narrative description of MR-specific bias
Firth 2020 ⁴⁷	Umbrella review	Modifiable health behaviours and major mental disorders	No	12/32	No	N/A	Yes/No	Statistical analysis results, use of sensitivity analysis and test for bidirectional effects	Yes
Frayling 2018 ⁴⁸	Systematic review	MR studies of T2D, coronary artery disease and hypertension	Yes	16/0	No	N/A	No	N/A	Yes
Grau-Perez 2019 ^{a, 22}	Systematic review	MR studies of environmentally responsive DNAm changes and the development of health states	Yes	15/0	Yes, Yes	Self-developed tool	No	N/A	Yes
Hu 2019 ⁴⁹	Systematic review	MR studies of atherosclerotic cardiovascular disease	Yes	58/0	No	N/A	No	N/A	Yes
Kei 2018 ⁵⁰	Systematic review	MR studies of serum uric acid levels and cardiovascular and renal disease risk	Yes	16/0	No	N/A	No	N/A	Yes
Kim 2020 ⁵¹	Umbrella review of systematic reviews and meta-analyses	Adiposity and cardiovascular disease events or mortality	No	27/11	No	N/A	Yes/No	Statistical power	Yes
Kohler 2018 ⁵²	Umbrella review of meta-analysis and MR studies	Environmental risk factors for depression	No	8/70	No	N/A	Yes/No	Proportion of variance in risk factors explained by genetic instruments	No

(Continued)

Table 3 Continued

Study ID	Type of article	Topic of the review	Were only MR studies included?	N of MR studies/N of non-MR studies	Risk-of-bias assessment in individual MR studies? If Yes, was a structured tool used?	Name and/or description of risk-of-bias assessment method	Evidence of causal effect assessment in individual MR studies? If Yes, was a structured method used?	Description of evidence of causal effect assessment method	Narrative description of MR-specific bias
Kuźma 2018 ^{a,24,53}	Systematic review	MR studies of risk factors and global cognitive function or dementia	Yes	18/0	Yes, Yes	Modified Q-Genie ²⁵	No	N/A	Yes
Li 2017 ⁵⁵	Umbrella review of systematic reviews and meta-analyses	Serum uric acid level and multiple health outcomes	No	36/101	No	N/A	Yes/No	Statistical significance of the effect estimate and statistical power	Yes
Lor 2019 ^{a,34}	Systematic review	MR analyses in oncological studies	Yes	77/0	No	N/A	No	N/A	Yes
Mamluk 2020 ^{a,28,56}	Systematic review	Maternal alcohol consumption in pregnancy and offspring outcomes at birth and later in life	No	9/14	Yes, Yes	Self-developed tool	No	N/A	Yes
Markozannes 2021 ⁵⁷	Umbrella review	C-reactive protein and health outcomes	No	37/55	Yes, No	Assessment of horizontal pleiotropy ^b	Yes/Yes	Statistical significance of the effect estimate	Yes
Meng 2019 ⁵⁸	Systematic review of MR studies and MR analysis	MR studies of vitamin D and health outcomes	Yes	65/0	No	N/A	No	N/A	No
Pearson-Stuttard 2021 ⁵⁹	Umbrella review	T2D and cancer incidence or mortality	No	8/20	Yes, No	Assessment of selection of genetic instrument	Yes/Yes	Statistical significance of the effect estimate	Yes
Pingault 2016 ⁶⁰	Systematic review	MR studies of psychopathology-related outcomes	Yes	19/0	No	N/A	No	N/A	Yes
Riaz 2018 ^{61,62}	Systematic review and meta-analysis of MR studies	MR studies of obesity and CVD	Yes	7/0	Yes, No	Evaluation of the three MR core assumptions	No	N/A	Yes
Robinson 2016 ⁶³	Literature review on MR methods, applications and limitations and systematic review of MR studies	MR studies of rheumatology	Yes	33/0	No	N/A	No	N/A	Yes

(Continued)

Table 3 Continued

Study ID	Type of article	Topic of the review	Were only MR studies included?	N of MR studies/N of non-MR studies	Risk-of-bias assessment in individual MR studies? If Yes, was a structured tool used?	Name and/or description of risk-of-bias assessment method	Evidence of causal effect assessment in individual MR studies? If Yes, was a structured method used?	Description of evidence of causal effect assessment method	Narrative description of MR-specific bias
Sommer 2018 ⁶⁴	Systematic review	Childhood and adolescent obesity and future cardiovascular morbidity and mortality later in life	No	1/85	No	N/A	No	N/A	No
Swerdlow 2016 ^{a,36}	Review on methods for selecting instruments for MR analysis and systematic review of MR studies	MR studies	Yes	231/0	No	N/A	No	N/A	Yes
Treur 2021 ^{a,31}	Systematic review	MR studies of poor mental health and substance use	Yes	63/0	Yes, Yes	Self-developed tool	No	N/A	Yes
Vasta 2021 ⁶⁵	Systematic review	Diabetes mellitus and amyotrophic lateral sclerosis	No	1/35	No	N/A	No	N/A	No
Yuan 2020 ⁶⁶	Systematic review and MR analysis	MR studies of risk factors of T2D	Yes	40/0	No	N/A	No	N/A	No
Zhang X 2019 ⁶⁷	Umbrella review	Non-genetic biomarkers and colorectal cancer	No	18/78	Yes, No	Assessment of horizontal pleiotropy	Yes/Yes	Statistical significance of the effect estimate, statistical power and evidence of bias due to directional pleiotropy	Yes
Zhang Z 2020 ⁶⁸	Systematic review	Vitamin D and non-alcoholic fatty liver disease	No	1/12	No	N/A	No	N/A	No

^aIncluded in the synthesis of tools for the assessing/evaluating MR studies.^bBased on the location of the SNPs. DNAm, DNA methylation; MR, Mendelian randomization; N/A, not applicable; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes.

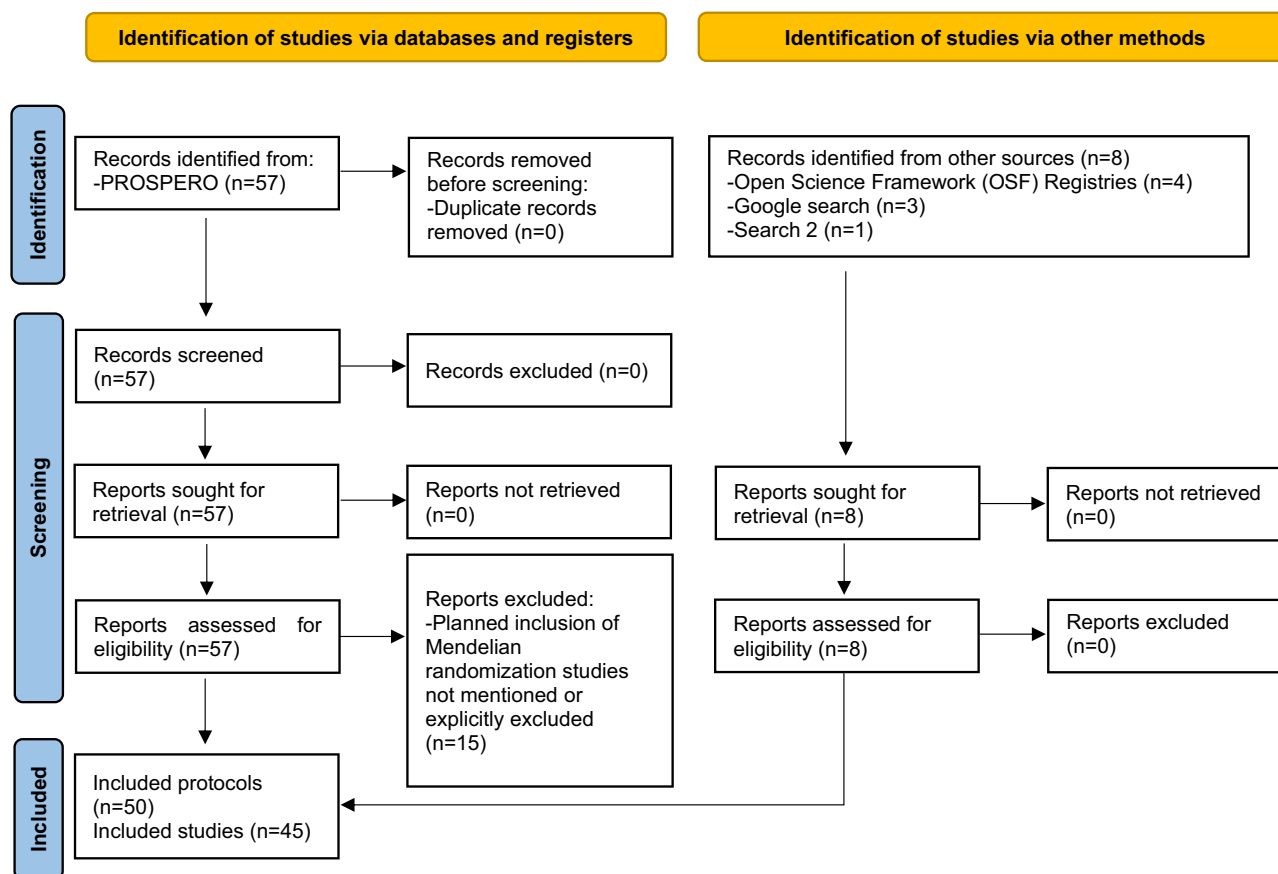


Figure 3 Flow diagram of identification, screening and inclusion of protocols of systematic reviews (and meta-analysis) planning to include Mendelian randomization studies

Supplementary data at *IJE* online). Five of the 45 included protocols were of published systematic reviews that were included in our sub-review of systematic reviews above.^{90,97,110,116,117} Of the 45 included protocols, 35 were for systematic reviews of primary studies and 10 were for umbrella reviews. Fifteen protocols were for reviews of MR studies only and 30 planned to include other study designs.

Eighteen protocols reported plans for a MR-specific risk-of-bias/quality-of-evidence assessment and 15 protocols reported plans for a non-MR-specific risk-of-bias/quality-of-evidence assessment. Of the 18 protocols with a MR-specific risk-of-bias/quality-of-evidence assessment, the use of a structured tool/method was planned in 11 protocols, the use of other methods/approaches was planned in 12 protocols and 1 protocol described the use of a method that the author planned to develop at the time of conducting the review. Of the 11 protocols describing use of a structured tool, Ibrahim⁸⁴ and Verdesen⁸⁵ planned to use STROBE-MR^{32,33} and other published literature, including the MR guidelines by Davies,²¹ LS Lee²⁶ planned to use a self-developed questionnaire (also included in our synthesis of tools) based on published guidelines including

Davies,²¹ Grover²⁷ and Burgess.²⁰ Markozannes⁹⁹ planned to use a self-developed tool based on the results of the main analysis and of the sensitivity analysis; Naassila^{100–102} planned to use Q-GENIE;²⁵ Shi^{105,106} planned to use a modified version of a recently developed tool (no reference provided); Visontay^{112,113} planned to use the tool developed by Mamluk;²⁸ and Wong¹¹⁵ planned to conduct risk-of-bias assessment based on the guidelines from Davies.²¹ Of the seven protocols describing a MR-specific risk-of-bias/quality-of-evidence assessment without using a structured tool, four planned an assessment based on the literature: Grover,^{80,81} Jiang⁸⁶ and van Oort¹¹¹ referred to the MR methods protocol published by Grover²⁷ and Julian⁸⁷ did not report any reference. Of the remaining four protocols, Saribaz¹⁰⁴ planned to develop a risk-of-bias assessment method at the time of conducting the review; M Lee⁹³ planned to perform a descriptive assessment of the MR methods and of the genetic variants used in included studies; Luo⁹⁶ planned to perform an assessment based on sensitivity analysis methods and different choices of genetic variants as instrumental variables; Treur¹¹⁰ planned to perform an assessment based on sensitivity analysis methods, on the choice of genetic variants, on the presence of sample

Table 4 List of included protocols of systematic reviews reporting Mendelian randomization studies Newcastle-Ottawa Scale²¹

Study ID	Topic of the review	Type of study	MR studies only?	Is there a plan to assess for risk of bias/quality of evidence in MR studies? If Yes, is a structured tool/ approach used?	What approach/method/tool?
Ansu 2020 ⁷⁰	Whole blood ionized magnesium in healthy adults	Systematic review	No	No	N/A
Baldwin 2020 ⁷¹	The impact of childhood maltreatment on mental health	Systematic review and meta-analysis	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale ⁶⁹
Cara 2020 ⁷²	Safety of enteral nutrition formulations with dietary fibre	Systematic review	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale ⁶⁹
Cheng 2019b, ¹¹⁷	Puberty timing and T2D	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Dack 2020 ⁷³	Early-life exposure to mercury, growth and neurodevelopment	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Desai 2021 ⁷⁴	Risk factors for dementia	Systematic review	Yes	Yes/Yes	Q-Genie ²⁵
Elsakloul 2016 ⁷⁵	Serum uric acid and cardiovascular diseases	Systematic review	No	NS/Yes	Pre-specified bespoke tool based on STROBE ¹¹⁸
Fan 2020 ⁷⁶	Habitual coffee consumption and lung function decline	Systematic review	No	NS/Yes	Tool for systematic reviews of observational studies that comprised four key domains: external validity, reporting, bias and confounding factors (no reference provided)
Fell 2020 ⁷⁷	Maternal smoking and orofacial clefts	Systematic review and meta-analysis	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Gianfredi 2019 ⁷⁸	Physical activity and depression	Systematic review	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale ⁶⁹
Gibson 2021 ⁷⁹	Reporting quality in MR studies using UK Biobank data	Systematic review of MR studies	Yes	No	N/A
Grover 2018 ^{80,81}	Risk factors for neurodegenerative diseases	Systematic review of MR studies	Yes	Yes/No	Assessment and reporting of MR studies based on previous published method protocol by Grover <i>et al.</i> ²⁷
Haan 2018 ^{82,83}	Alcohol, tobacco and caffeine consumption in pregnancy and externalizing disorders in offspring	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Ibrahim 2020 ⁸⁴	MR studies of abdominal aortic aneurysms	Systematic review of MR studies and meta-analysis	Yes	Yes/Yes	STROBE-MR ^{23,85} and other publications
Jiang 2019 ⁸⁶	Causal factors associated with risk or survival in lung cancer	Systematic review of MR studies	Yes	Yes/No	Assessment of risk of bias and quality of reporting of MR studies based on previous method protocol by Grover <i>et al.</i> ²⁷ . Assessment of the robustness and credibility of the data synthesis using sensitivity analysis

(Continued)

Table 4 Continued

Study ID	Topic of the review	Type of study	MR studies only?	Is there a plan to assess for risk of bias/quality of evidence in MR studies? If Yes, is a structured tool/ approach used?	What approach/method/tool?
Julian 2020 ⁸⁷	MR studies of neurodegenerative disease	Systematic review of MR studies	Yes	Yes/No	Assessment of risk of bias, evidence base for methodological strengths and weaknesses using the published literature (no reference provided)
Karwatowska 2020 ^{88,89}	Risk factors for disruptive behaviours	Systematic review and meta-analysis	No	NS/Yes	Adapted version of the ROBINS-I checklist ¹⁹
Kim 2020a ⁹⁰	Obesity and cardiovascular outcomes	Umbrella review	No	No	N/A
Kim 2020b ⁹¹	Obesity and gastroenterological diseases	Umbrella review	No	No	N/A
Kim 2021 ^{b,92}	Obesity and renal and genitourinary outcomes	Umbrella review	No	No	N/A
Lee LS 2020 ^{a,26}	Risk factors for dementia	Systematic review of MR studies and meta-analysis	Yes	Yes/No	Assessment of quality using a self-developed questionnaire based on published guidelines
Lee M 2018 ⁹³	MR studies using adiposity as an exposure	Systematic review of MR studies	Yes	Yes/Yes	Descriptive assessment of choice of methods and genetic variants used in included studies
Lemus 2021 ⁹⁴	T2D and incidence of 17 types of cancer	Systematic review and meta-analysis	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Liu 2020 ⁹⁵	Risk factors for coronavirus disease 19 (COVID-19)	Umbrella review	No	No	N/A
Luo 2017 ⁹⁶	MR studies compared with randomized-controlled trials	Systematic review	No	Yes/No	Assessment of the robustness and credibility of an estimate based on sensitivity analysis methods and different choices of genetic variants as instrumental variables
Mamluk 2015 ^{b,97}	Prenatal alcohol exposure on pregnancy and childhood outcomes	Systematic review	No	NS/No	Assessment of quality of evidence based on whether studies have adjusted for smoking and maternal education/social class as potential confounders in their final model
Maretzke 2018 ^{c,98}	Role of vitamin D in preventing and treating selected extra-skeletal diseases	Umbrella review	No	No	N/A
Markozannes 2021 ⁹⁹	Genetically predicted risk factors associated with cancer risk	Systematic review of MR studies	Yes	Yes/Yes	Self-developed tool based on the results of the main analysis and of the sensitivity analysis
Naassila 2021 ¹⁰⁰	Alcohol intake and risk of cardiovascular diseases	Systematic review	No	Yes/Yes	Q-Genie ²⁵
Naassila 2021 ¹⁰¹	Alcohol intake and risk of neurological diseases	Systematic review	No	Yes/Yes	Q-Genie ²⁵

(Continued)

Table 4 Continued

Study ID	Topic of the review	Type of study	MR studies only?	Is there a plan to assess for risk of bias/quality of evidence in MR studies? If Yes, is a structured tool/ approach used?	What approach/method/tool?
Naassila 2021 ¹⁰²	Alcohol intake and cancers, neurological, cardiovascular and liver diseases	Systematic review	Yes	Yes/Yes	Q-Genie ²⁵
Romo 2018 ¹⁰³	Conduct and reporting of MR studies	Systematic review of MR studies	Yes	No	N/A
Saribaz 2020 ¹⁰⁴	Environmental risk factors of child and adolescents' depressive and anxious psychopathology	Systematic review	No	Yes/NR	Self-developed method developed at the time of review
Shi 2020 ^{105,106}	Prenatal alcohol exposure and offspring health outcomes	Umbrella review	No	Yes/Yes	Modified recently developed tool (reference not provided)
Solmi 2018a ¹⁰⁷	Safety and efficacy of cannabinoids and cannabis in treating medical conditions	Umbrella review	No	No	N/A
Solmi 2018b ¹⁰⁸	Psychosis and non-communicable general medical conditions	Umbrella review	No	No	N/A
Suh 2021 ¹⁰⁹	Risk factors for cardiovascular multimorbidity	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale ⁶⁹
Treur 2019 ^{b,110}	Substance use, cognitive functioning and psychiatric disorders	Systematic review of MR studies	Yes	Yes/No	Descriptive assessment based on MR study design, choice of genetic variants, whether there was sample overlap in the case of two-sample MR studies and the use of sensitivity analyses
van Oort 2020 ¹¹¹	Alcohol consumption and its causal relationship with mortality, cardiometabolic diseases and risk factors	Systematic review of MR studies	Yes	Yes/No	Assessment of the quality of MR studies based on previous published method protocol by Grover <i>et al.</i> ²⁷ with focus on MR design, the quality of the genetic instrument and the validation of the MR assumptions
Verdiesen 2021 ⁸⁵	Causal risk factors for breast cancer	Systematic review of MR studies and meta-analysis	Yes	Yes/Yes	STROBE-MR and a published checklist by Davies <i>et al.</i> ²¹
Visontay 2020 ^{112,113}	Alcohol consumption and health outcomes	Systematic review	No	Yes/Yes	Recently developed risk-of-bias tools specific to MR studies, natural experiments, and other genetic-based methods by Mamluk <i>et al.</i> ²⁸
Wang 2018 ¹¹⁴	Vitamin D deficiency as a causal risk factor	Umbrella review	No	NS/Yes	Assessment of risk of bias as described in the Cochrane risk-of-bias tool (reference not provided)

(Continued)

Table 4 Continued

Study ID	Topic of the review	Type of study	MR studies only?	Is there a plan to assess for risk of bias/quality of evidence in MR studies? If Yes, is a structured tool/approach used?	What approach/method/tool?
Wong 2021 ¹¹⁵	Factors contributing to higher coronavirus disease 19 (COVID-19) risk or its severity	Living systematic review	Yes	Yes/Yes	Assessment of risk of bias based on a published checklist by Davies <i>et al.</i> ²¹
Yan 2020 ⁵⁴	Metabolomic profiling of amino acids in serum/plasma and urine and risk of cardiovascular disease and T2D	Systematic review and meta-analysis	No	NS/Yes	Assessment of risk of bias using the Cochrane risk-of-bias tool (for randomized-controlled trials) ¹⁸ and the ROBINS-I ¹⁹
Zhang 2018 ^{b,116}	Non-genetic biomarkers and risk of colorectal cancer	Umbrella review	No	No	N/A

^aIncluded in the synthesis of tools for assessing, conducting and reporting MR studies.

^bProtocols of published systematic reviews included in this article.

^cProtocol of published systematic review not included in this article. MR, Mendelian randomization; N/A, not applicable; NR, not reported; NS, non-specifically; ROBINS-I, risk of bias in non-randomized studies of intervention; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; T2D, type 2 diabetes.

overlap (two-sample MR studies) and on the use of sensitivity analyses.

Of the 15 protocols in which a non-MR-specific risk-of-bias assessment is reported, 14 used structural tools and Mamluk⁹⁷ planned to assess risk of bias on whether adjustment for potentially relevant confounders was conducted. Of the 14 structured tools used for non-MR-specific risk-of-bias assessment, Cheng,¹¹⁷ Dack,⁷³ Fell,⁷⁷ Haan,^{82,83} Lemus⁹⁴ and Suh¹⁰⁹ planned to use NOS⁶⁹ and Baldwin,⁷¹ Cara^{72,109} and Gianfredi⁷⁸ planned to use a modified version of NOS; Elsakloul⁷⁵ planned to use STROBE;¹¹⁸ Fan⁷⁶ planned to use a quality-assessment tool for systematic reviews of observational studies that comprised external validity, reporting, bias and confounding factors, but a reference was not provided; Karwatowska^{88,89} planned to use ROBINS-I;¹⁹ Yan⁵⁴ planned to use the ROB-2¹⁸ and the ROBINS-I¹⁹ tools; and Wang¹¹⁴ planned to use the Cochrane risk-of-bias assessment tool (no details provided).

Discussion

Our systematic review of tools identified 14 instruments developed for the evaluation, conduct and/or reporting of MR studies. Half of the tools were designed (or used) either entirely or partially for the evaluation of MR studies. Most of these tools were developed for application within a systematic review,^{22,24,26,28,31} whereas only two were developed for general use.^{20,21} Despite notable variability in the structure and content of the evaluating tools, all tools contained items addressing the validity of the three core IV assumptions. In addition, all but one of the tools addressed bias related to the selection of the population(s) or sample(s), including population heterogeneity, sample overlap, choice of controls and selection bias. Just over half of the tools addressed bias related to the genetic instrument, including linkage disequilibrium, construct of the genetic score and lack of variants harmonization, and addressed the conduct of sensitivity analysis. Fewer than half of the evaluating tools addressed bias due to measurement errors and only one tool addressed bias due to other sources including missing data. Although it was not in our scope to critically appraise the identified tools, by compiling a list and inspecting the content of these tools we found that all tools, including these designed for reporting and conducting, addressed these assumptions or conditions within the MR analysis that, when violated, lead to potential bias of the MR causal estimate. Of the seven tools designed (or used) for evaluation of MR studies, three tools included a scoring/rating system^{24,28,31} but none of the tools attempted to predict the likely direction of bias

(i.e. whether the results are biased away from or towards the null).

Consistently with the lack of formal tools for assessment of risk of bias in MR studies, only a small proportion (26%) of the systematic reviews of MR studies included in our review conducted a risk-of-bias assessment and only 23% of the included reviews conducted an assessment of evidence of causal effect within individual MR studies. Nevertheless, most of the reviews included a narrative description of MR-related bias and limitations (74%) and—as observed in the content of the tools—among these, most of the reviews addressed bias related to the core IV assumptions of relevance (IV1) and exclusion restriction (IV3) (71% and 86%, respectively), but only 57% addressed bias related to the independence assumption (IV2), whereas 61% addressed bias related to the genetic instrument and only 21% addressed bias related to the selection of the population or sample.

In contrast with published systematic reviews, when we looked at protocols of systematic reviews of (or including) MR studies, a plan to conduct an assessment was reported in 73% of the protocols included in our reviews, although only in 40% was the approach or methodology used specific for MR studies. This higher proportion may reflect an increased focus on risk of bias over time or may reflect a tendency for review teams who publish their protocols to include risk-of-bias assessments in their plans. Of protocols that specified methodologies specific to MR studies, only 39% planned to use a structured tool, including the STROBE-MR,^{32,33} Q-GENIE,²⁵ a self-developed tool included in our synthesis of tools²⁶ and a tool developed within another systematic review.²⁸ One review protocol planned to use a recently developed tool that, similarly to the tool developed by Mamluk,²⁸ consisted of five questions, one for bias domain, including instrument bias, genetic confounding and selection bias. The rest of the protocols not planning to use a structured tool proposed other informal ways to address bias, including assessment based on the validation of the three IV core assumptions, the choice of genetic instruments, the use of sensitivity analysis and description of MR analysis design, and some of these approaches were based on MR literature including MR guidelines by Davies²¹ and Grover.²⁷

Our review has strengths and limitations. First, we included published and unpublished articles by searching several relevant databases for peer-reviewed articles, preprints archives and Google Scholar for preprints articles and unpublished studies. Furthermore, we developed specific search strings for each objective with the assistance of an information specialist. However, as some of the tools we have identified were developed within other types of articles, including literature reviews and systematic reviews

of MR and non-MR studies, it is possible that our searches may have missed some tools. As data extraction was performed by a single author, it is possible that some errors in data collection were made. Our classification of items into bias domains and specific issues is to an extent arbitrary and some items could have been classified in accordance with more than one bias or limitation. For example, we classified linkage disequilibrium as relevant to the choice of genetic variant because it mainly introduces horizontal pleiotropy¹⁷ (IV3 domain) although it has been argued to be associated also with confounding (IV2 domain).⁵

By summarizing the currently available knowledge on methods and approaches for assessment of risk of bias in MR studies, our longer-term aim was to identify potential items for inclusion in a structured tool for risk-of-bias assessment in MR studies. We are not able to make a recommendation on what tool(s) should be adopted to assess MR studies, as none of the tools identified by our searches appears to have been formally tested or validated. A systematic process to test reliability using formal studies of agreement should be conducted before the tools can be recommended for general use. Validation studies could in theory be undertaken using a meta-epidemiological approach, in which effect estimates are compared between studies with different bias-related features. However, such studies require large numbers of meta-analyses with their included studies all having been assessed using the same tool(s).

Nevertheless, the content of the tools that we have identified in our review will be a useful source of information on what bias/limitations reviewers should be aware of when conducting a systematic review (and meta-analysis) including results from MR studies. This suggests that issues to address include those arising from departures from the IV assumptions, those related to the choice of genetic instrument(s) and those arising from the population from which the data are collected (particularly in two-sample MR), in addition to more traditional non-MR-specific epidemiological biases.

Ethics approval

Not applicable. All the work was developed using published data.

Data availability

All materials used in this study are available in the [Supplementary material](#), available as [Supplementary data](#) at *IJE* online, or the main text.

Supplementary data

[Supplementary data](#) are available at *IJE* online.

Author contributions

F.S. is the guarantor of the review. F.S. led the writing of the manuscript and developed the idea with J.P.T.H. and M.R.M.; F.S. and S.D. developed the search strategies; F.S. and M.G. conducted the study selection; F.S. extracted and analysed the data; F.S. drafted the manuscript with critical feedback and revisions from M.G., S.D., K.T., G.D.S., M.R.M. and J.P.T.H. All authors approved the final version of the manuscript. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This work was supported by a Cancer Research UK programme grant (the Integrative Cancer Epidemiology Programme C18281/A29019 to F.S. and M.R.M.). G.D.S., K.T. and M.R.M. are part of the Medical Research Council Integrative Epidemiology Unit (MRC IEU) at the University of Bristol (MC_UU_00011/1 to G.D.S.; MM_UU_00011/3 to K.T.; MM_UU_00011/7 to M.R.M.). J.P.T.H. is a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0617-10145) and is supported by NIHR Bristol Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol and by the NIHR Applied Research Collaboration West (ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust.

Conflict of interest

None declared. For the sake of transparency, we acknowledge that G.D.S, M.R. M and J.P.T.H. are authors of tools included in this review. None of the authors received financial compensation for this work.

References

- Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
- Sanderson E, Glymour M, Holmes MV *et al.* Mendelian randomization. *Nat Rev Methods Primers* 2022;2:7.
- Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;4:e352.
- Davey Smith G, Holmes MV, Davies NM, Ebrahim S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur J Epidemiol* 2020;35:99–111.
- Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63.
- Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33:30–42.
- Labrecque J, Swanson SA. Understanding the assumptions underlying instrumental variable analyses: a brief review of falsification strategies and related tools. *Curr Epidemiol Rep* 2018;5:214–20.
- Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29:722–29.
- Davey Smith G. Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice? *Int J Epidemiol* 2012;41:236–47.
- Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* 2016;45:1717–26.
- Zhao Q, Wang J, Spiller W, Bowden J, Small DS. Two-sample instrumental variable analyses using heterogeneous samples. *Statist Sci* 2019;34:317–33.
- Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011;30:1312–23.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–98.
- Brumpton B, Sanderson E, Heilbron K *et al.*; 23andMe Research Team. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* 2020;11:3519.
- Lawlor D, Richmond R, Warrington N *et al.* Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: sources of bias and methods for assessing them [version 1; peer review: 4 approved]. *Wellcome Open Res* 2017;2:11.
- Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016;40:597–608.
- Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018;27:R195–R208.
- Sterne JAC, Savovic J, Page MJ *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898.
- Sterne JA, Hernan MA, Reeves BC *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- Burgess S, Davey Smith G, Davies NM *et al.* Guidelines for performing Mendelian randomization investigations [version 2; peer review: 2 approved]. *Wellcome Open Res* 2019;4:186.
- Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362:k601.
- Grau-Perez M, Agha G, Pang Y, Bermudez JD, Tellez-Plaza M. Mendelian randomization and the environmental epigenetics of health: a systematic review. *Curr Environ Health Rep* 2019;6:38–51.
- 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;44:496–511.
- Kuzma E, Hannon E, Zhou A *et al.* Which risk factors causally influence dementia? A systematic review of Mendelian randomization studies. *J Alzheimers Dis* 2018;64:181–93.
- Sohani ZN, Meyre D, de Souza RJ *et al.* Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. *BMC Genet* 2015;16:50.

26. Lee LS, Whiteley W, Walker R. Systematic review and meta-analysis of Mendelian randomisation studies on modifiable risk factors for dementia. *protocols.io*. <https://dx.doi.org/10.17504/protocols.io.bpeemjbe> (23 June 2022, date last accessed).
27. Grover S, Del Greco MF, Stein CM, Ziegler A. Mendelian randomization. In: Elston RC (ed). *Statistical Human Genetics: Methods and Protocols*. New York, NY: Springer New York, 2017; 581–628.
28. Mamluk L, Jones T, Ijaz S *et al*. Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *Int J Epidemiol* 2021;**49**:1972–95.
29. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012; **175**:332–39.
30. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafò MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014;**13**:99–106.
31. Treur JL, Munafo MR, Logtenberg E, Wiers RW, Verweij KJH. Using Mendelian randomization analysis to better understand the relationship between mental health and substance use: a systematic review. *Psychol Med* 2021;**51**: 1593–24.
32. Davey Smith G, Davies NM, Dimou N *et al*. STROBE-MR: guidelines for strengthening the reporting of Mendelian randomization studies. *PeerJ Preprints* 2019;**7**:e27857v1.
33. Skrivankova VW, Richmond RC, Woolf BAR *et al*. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA* 2021;**326**:1614–21.
34. Lor GCY, Risch HA, Fung WT *et al*. Reporting and guidelines for Mendelian randomization analysis: a systematic review of oncological studies. *Cancer Epidemiol* 2019;**62**:101577.
35. Lawlor DA, Wade K, Borges MC, *et al*. A Mendelian Randomization dictionary: useful definitions and descriptions for undertaking, understanding and interpreting Mendelian randomization studies. OSF Preprints. doi:10.31219/osf.io/6yzz7, 5 February 2019, preprint: not peer reviewed.
36. Swerdlow DI, Kuchenbaecker KB, Shah S *et al*. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 2016;**45**:1600–16.
37. Abbasi A. Mendelian randomization studies of biomarkers and type 2 diabetes. *Endocr Connect* 2015;**4**:249–60.
38. Abbasi A, Sahlqvist A-S, Lotta L *et al*. A systematic review of biomarkers and risk of incident type 2 diabetes: an overview of epidemiological, prediction and aetiological research literature. *PLoS One* 2016;**11**:e0163721.
39. Bellou L, Bellou V, Evangelou E, Tzoulaki I. Environmental factors and risk of multiple sclerosis: findings from meta-analyses and Mendelian randomization studies. *Mult Scler* 2020;**26**:397–404.
40. Bellou L, Kohler CA, Stefanis N *et al*. Risk factors and peripheral biomarkers for schizophrenia spectrum disorders: an umbrella review of meta-analyses. *Acta Psychiatr Scand* 2018;**137**: 88–97.
41. Bellou V, Bellou L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: an exposure-wide umbrella review of meta-analyses. *PLoS ONE* 2018;**13**:e0194127.
42. Bergmans RS, Rapp A, Kelly KM, Weiss D, Mezuk B. Understanding the relationship between type 2 diabetes and depression: lessons from genetically informative study designs. *Diabet Med* 2021;**38**:e14399.
43. Bochud M, Rousson V. Usefulness of Mendelian randomization in observational epidemiology. *Int J Environ Res Public Health* 2010;**7**:711–28.
44. Carnegie R, Zheng J, Sallis HM *et al*. Mendelian randomisation for nutritional psychiatry. *Lancet Psychiatry* 2020;**7**:208–16.
45. Cheng TS, Day FR, Lakshman R, Ong KK. Association of puberty timing with type 2 diabetes: a systematic review and meta-analysis. *PLoS Med* 2020;**17**:e1003017.
46. 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* 2021;**35**:130–42.
47. Firth J, Solmi M, Wootton RE *et al*. A meta-review of ‘lifestyle psychiatry’: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry* 2020;**19**:360–80.
48. Frayling TM, Stoneman CE. Mendelian randomisation in type 2 diabetes and coronary artery disease. *Curr Opin Genet Dev* 2018;**50**:111–20.
49. Hu Q, Hao P, Liu Q *et al*. Mendelian randomization studies on atherosclerotic cardiovascular disease: evidence and limitations. *Sci China Life Sci* 2019;**62**:758–70.
50. Kei A, Koutsouka F, Makri A, Elisaf M. Uric acid and cardiovascular risk: what genes can say. *Int J Clin Pract* 2018;**72**: e13048.
51. Kim MS, Kim WJ, Khera AV, Won H-H. Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis. medRxiv; <https://doi.org/10.1101/2020.08.18.20176578>, 22 August 2020, preprint: not peer reviewed.
52. Kohler CA, Evangelou E, Stubbs B *et al*. Mapping risk factors for depression across the lifespan: an umbrella review of evidence from meta-analyses and Mendelian randomization studies. *J Psychiatr Res* 2018;**103**:189–207.
53. Kuzma E, Hannon E, Zhou A *et al*. A systematic review of Mendelian randomization studies investigating causal associations between risk factors and dementia. *Alzheimers Dement* 2017;**13**:P1180.
54. Yan P, Zhao J, Li M. Metabolomic profiling of amino acids and risk of cardiovascular disease and type 2 diabetes: a systematic review and meta-analysis. PROSPERO 2021 CRD42021199408. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021199408.
55. Li X, Meng X, Timofeeva M *et al*. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ (Clinical Research Ed)* 2017;**357**:j2376.
56. Mamluk L, Edwards HB, Savovic J *et al*. Prenatal alcohol exposure and pregnancy and childhood outcomes: a systematic review of alternative analytical approaches. *Lancet* 2016;**388**:73.

57. Markozannes G, Koutsioumpa C, Cividini S *et al*. Global assessment of C-reactive protein and health-related outcomes: an umbrella review of evidence from observational studies and Mendelian randomization studies. *Eur J Epidemiol* 2021;36: 11–36.
58. Meng X, Li X, Timofeeva MN *et al*. Phenome-wide Mendelian-randomization study of genetically determined vitamin D on multiple health outcomes using the UK Biobank study. *Int J Epidemiol* 2019;48:1425–34.
59. Pearson-Stuttard J, Papadimitriou N, Markozannes G *et al*. Type 2 diabetes and cancer: an umbrella review of observational and Mendelian randomisation studies. *Cancer Epidemiol Biomarkers Prev* 2021;30:1218–28.
60. Pingault J-B, Cecil CAM, Murray J, Munafò MR, Viding E. Causal inference in psychopathology: a systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology. *Psychopathol Rev* 2017;a4:4–25.
61. Riaz H, Khan MS, Siddiqi TJ *et al*. Association between obesity and cardiovascular outcomes: a systematic review and meta-analysis of Mendelian randomization studies. *JAMA Netw Open* 2018;1:e183788.
62. Riaz H, Shah N, Khan MS, Goyal A, Siddiqi TJ, Ahmed H. Is obesity causally linked with cardiovascular outcomes? A meta-analysis of Mendelian randomization studies. *JACC* 2018;71: A1889.
63. Robinson PC, Choi HK, Do R, Merriman TR. Insight into rheumatological cause and effect through the use of Mendelian randomization. *Nat Rev Rheumatol* 2016;12:486–96.
64. Sommer A, Twig G. The impact of childhood and adolescent obesity on cardiovascular risk in adulthood: a systematic review. *Curr Diab Rep* 2018;18:91.
65. Vasta R, D'Ovidio F, Logroscino G, Chio A. The links between diabetes mellitus and amyotrophic lateral sclerosis. *Neurol Sci* 2021;42:1377–87.
66. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia* 2020;63:2359–71.
67. Zhang X, Gill D, He Y *et al*. Non-genetic biomarkers and colorectal cancer risk: umbrella review and evidence triangulation. *Cancer Med* 2020;9:4823–35.
68. Zhang Z, Thorne JL, Moore JB. Vitamin D and nonalcoholic fatty liver disease. *Curr Opin Clin Nutr Metab Care* 2019;22: 449–58.
69. Wells G, Shea B, O'Connell D, *et al*. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2013. <http://www.ohrica.org/programs/clinical-epidemiology/oxford.asp> (23 June 2022, date last accessed).
70. Ansu V, Wallace T, Cara K, Spence L, Chung M, Gletsu-Miller N. Systematic review and meta-analysis to determine a reference range for ionized magnesium. PROSPERO 2020 CRD42020216100. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020216100 (23 June 2022, date last accessed).
71. Baldwin J, Wang B, Karwatowska L, Schoeler T, Munafò M, Pingault JB. The impact of childhood maltreatment on mental health: a meta-analysis of quasi-experimental studies. PROSPERO 2020 CRD42020187520. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020187520 (23 June 2022, date last accessed).
72. Cara KC, Beauchesne AR, Chung M, Wallace T. Systematic review and meta-analysis on the safety of using enteral nutrition formulations containing dietary fiber in hospitalized critical care patients. PROSPERO 2020 CRD42020204181. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020204181 (23 June 2022, date last accessed).
73. Dack K, Fell M, Lewis S, Taylor C, Havdahl A. The effects of early life exposure to mercury on growth and neurodevelopment: a systematic review. PROSPERO 2020 CRD42020221146. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020221146 (23 June 2022, date last accessed).
74. Desai R, Stott J. Risk factors and dementia: systematic review of Mendelian randomization studies. PROSPERO 2021 CRD42021254793. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021254793 (23 June 2022, date last accessed).
75. Elsakloul F, Dawson J. The genetic determinants of serum uric acid and their association with cardiovascular diseases: a systematic review. PROSPERO 2016 CRD42016037362. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016037362 (23 June 2022, date last accessed).
76. Fan H, Chen Y, Lee YL. Habitual coffee consumption and lung function decline: multi-country meta-analyses and Mendelian randomization studies. PROSPERO 2020 CRD42020184742. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020184742 (23 June 2022, date last accessed).
77. Fell M, Dack K, Chummun S, Wren Y, Lewis S. The role of maternal cigarette smoking in the aetiology of cleft lip and palate: a systematic review and meta-analysis. PROSPERO 2020 CRD42020222837. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020222837 (23 June 2022, date last accessed).
78. Gianfredi V, Blandi L, Cacitti S, Minelli M. Physical activity and depression: a systematic review and meta-analysis on the association between patterns of objectively measured physical activity and risk of depression in adults. PROSPERO 2020 CRD42020132860. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020132860 (23 June 2022, date last accessed).
79. Gibson M, Spiga F, Richmond R, Munafò MR. A systematic review of the reporting quality of studies that use Mendelian randomisation in UK Biobank. OSF 2021. <https://doi.org/10.17605/OSF.IO/HN7WD> (23 June 2022, date last accessed).
80. Grover S, Del Greco F, König IM. Identification of causal risk factors for neurodegenerative diseases using a Mendelian randomization approach. PROSPERO 2018 CRD42018091434. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018091434 (23 June 2022, date last accessed).
81. Grover S, Del Greco F, König IM. Evaluating the current state of Mendelian randomization studies: a protocol for a systematic review on methodological and clinical aspects using neurodegenerative disorders as outcome. *Syst Rev* 2018;7:145.
82. Haan E, Schellhas L, Westmoreland K, Sallis H, Taylor G, Munafò M. Associations between alcohol, tobacco and caffeine consumption in pregnancy and externalising disorders in offspring: a systematic review and meta-analysis. PROSPERO

- 2018 CRD42018094810. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018094810 (23 June 2022, date last accessed).
83. Haan E, Schellhas L, Sallis H, Taylor G, Munafo MR. Associations between alcohol, tobacco and caffeine consumption in pregnancy and externalising disorders in offspring: a systematic review and meta-analysis. *OSF*. 2019. <https://doi.org/10.17605/OSF.IO/D9WZK> (23 June 2022, date last accessed).
 84. Ibrahim M, Golledge J, Singh T, Thanigaimani S, Morris D. Systematic review and meta-analysis of Mendelian randomisation of abdominal aortic aneurysms. *PROSPERO* 2020 CRD42020203479. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020203479 (23 June 2022, date last accessed).
 85. Verdiesen R, Schmidt M. Causal risk factors for breast cancer: a systematic review and meta-analysis of Mendelian randomization studies. *PROSPERO* 2021 CRD42021239009. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021239009 (23 June 2022, date last accessed).
 86. Jiang M, Xu W, Liu G. A systematic review of Mendelian randomization studies in lung cancer. *PROSPERO* 2019 CRD42019135756. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019135756 (23 June 2022, date last accessed).
 87. Julian T, Cooper-Knock J, Shaw P. Mendelian randomisation in neurodegenerative disease: a systematic review. *PROSPERO* 2020 CRD42020221080. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020221080 (23 June 2022, date last accessed).
 88. Karwowska L, Russell S, Solmi F *et al*. Risk factors for disruptive behaviours: protocol for a systematic review and meta-analysis of quasi-experimental evidence. *BMJ Open* 2020;10:e038258.
 89. Karwowska L, Russell S, De Stavola B, Jaffee S, Pingault JB, Viding E. Risk factors for disruptive behaviours: a systematic review and meta-analysis of quasi-experimental evidence. *PROSPERO* 2020 CRD42020169313. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020169313 (23 June 2022, date last accessed).
 90. Kim MS, Jun KW. Obesity as a risk factor for cardiovascular outcomes: umbrella review of the literature. *PROSPERO* 2020 CRD42020179469. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020179469 (23 June 2022, date last accessed).
 91. Kim MS, Lee I, Park S. Impact of obesity on gastroenterological diseases: umbrella review of systematic review and meta-analysis. *PROSPERO* 2020 CRD42020209826. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020209826 (23 June 2022, date last accessed).
 92. Kim MS, Park S. Impact of obesity on renal and genitourinary outcomes: an umbrella review of systematic reviews and meta-analyses. *PROSPERO* 2021 CRD42021244692. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021244692 (23 June 2022, date last accessed).
 93. Lee M, McGuinness LAA. Systematic review of Mendelian randomization studies using adiposity as an exposure. *PROSPERO* 2018 CRD42018096684. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018096684 (23 June 2022, date last accessed).
 94. Lemus F, Hawkins N, Boyd KA. Type 2 diabetes as a driving factor in the incidence of 17 types of cancer: a systematic review and meta-analysis. *PROSPERO* 2021 CRD42021257479. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021257479 (23 June 2022, date last accessed).
 95. Liu D, Tian Q, Cao M, *et al*. Mapping risk factors for COVID-19: an umbrella review of evidence from observational studies, and Mendelian randomization studies. *PROSPERO* 2020 CRD42020205289. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020205289 (23 June 2022, date last accessed).
 96. Luo S, Zhao J, Yang Q, Schooling M. Mendelian randomization studies compared to randomized controlled trials: a systematic review and meta-analysis. *PROSPERO* 2017 CRD42017073731. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017073731 (23 June 2022, date last accessed).
 97. Mamluk L, Zuccolo L, Moore T, Richards A. Systematic review of the effects of low-moderate prenatal alcohol exposure on pregnancy and childhood outcomes. *PROSPERO* 2015 CRD42015015941. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42015015941 (23 June 2022, date last accessed).
 98. Maretzke F, Bechthold A, Egert S, *et al*. Role of vitamin D in preventing and treating selected extraskelatal diseases: an umbrella review. *PROSPERO* 2019 CRD42019103670. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019103670 (23 June 2022, date last accessed).
 99. Markozannes G, Kannelopoulou A, Dimopoulou O, *et al*. Systematic review of Mendelian randomization studies on risk of cancer. *OSF*. 31 March. <https://doi.org/10.17605/OSF.IO/2RUCT> (23 June 2022, date last accessed).
 100. Naassila M, Naouras B, Aubin HJ, *et al*. Effects of alcohol intake on the risk of cardiovascular diseases: systematic review. *PROSPERO* 2021 CRD42021252410. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021252410 (23 June 2022, date last accessed).
 101. Naassila M, Naouras B, Aubin HJ, *et al*. Effects of alcohol intake on the risk of neurological diseases: systematic review. *PROSPERO* 2021 CRD42021251777. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021251777 (23 June 2022, date last accessed).
 102. Naassila M, Bouajila N, Aubin HJ, *et al*. Alcohol intake and cancers, neurological, cardiovascular and liver diseases: systematic review implementing Mendelian randomization studies. *PROSPERO* 2021 CRD42021246154. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021246154 (23 June 2022, date last accessed).
 103. Romo MV, Huang JV, Lun S, Yeung A, Zhao J, Schooling MC. The conduct and reporting of Mendelian randomization studies: a systematic review. *PROSPERO* 2018 CRD42018087938. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018087938 (23 June 2022, date last accessed).
 104. Saribaz ZE, Rowe R, Debowska A, Hales G. Environmental risk factors of child and adolescents' depressive and anxious psychopathology: a systematic review of causally informative

- study designs. PROSPERO 2021 CRD42021224578. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021224578 (23 June 2022, date last accessed).
105. Shi X, Rogne T, Liew Z, Nyhan K, Bracken M, Wallach J. Prenatal alcohol exposure and offspring health outcomes: an umbrella review of evidence from observational and Mendelian randomization studies. PROSPERO 2020 CRD42020215670. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020215670 (23 June 2022, date last accessed).
106. Shi X, Rogne T, Liew Z, Nyhan K, Bracken MB, Wallach JD. Prenatal alcohol exposure and offspring health outcomes: an umbrella review of evidence from observational and Mendelian randomization studies. OSF, 2020. <https://osf.io/8rtph> (23 June 2022, date last accessed).
107. Solmi M, Stubbs B, Miola A, *et al.* Cannabinoids and cannabis: useful or harmful? An umbrella review of meta-analyses of randomized controlled trials and observational studies. PROSPERO 2018 CRD42018093045. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018093045 (23 June 2022, date last accessed).
108. Solmi M, Papola D, Barbui C, *et al.* Associations of psychosis and non-communicable general medical conditions: an umbrella review of observational and Mendelian randomization studies. PROSPERO 2018 CRD42018111983. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018111983 (23 June 2022, date last accessed).
109. Suh JW, Sun S, Chadni MJ, *et al.* Risk factors for cardiovascular multimorbidity: a systematic review of prospective studies. PROSPERO 2021 CRD42021240491. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021240491 (23 June 2022, date last accessed).
110. Treur J, Logtenberg E, Munafò M, Wiers R, Verweij K. Using Mendelian randomization to unravel the complex relationship of substance use with cognitive functioning and psychiatric disorders: a systematic review. PROSPERO 2019 CRD42019133182. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019133182 (23 June 2022, date last accessed).
111. van Oort S, van de Luitgaarden I, Beulens J, *et al.* Alcohol consumption and its causal relationship with mortality, cardio-metabolic diseases, and risk factors: a systematic review of mendelian randomization studies. PROSPERO 2020 CRD42020151510. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020151510 (23 June 2022, date last accessed).
112. Visontay R, Sunderland M, Slade T, Wilson J, Mewton L. Are there non-linear relationships between alcohol consumption and long-term health? A systematic review of observational studies employing approaches to improve causal inference. PROSPERO 2020 CRD42020185861. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020185861 (23 June 2022, date last accessed).
113. Visontay R, Sunderland M, Slade T, Wilson J, Mewton L. Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ Open* 2021;11:e043985.
114. Wang Y, Liu D. Mendelian randomization studies using vitamin D deficiency as a causal risk factor: umbrella review of systematic reviews and meta-analyses of observational studies, randomised controlled trials, and Mendelian randomization studies. PROSPERO 2019 CRD42019115878. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019115878 (23 June 2022, date last accessed).
115. Wong THT, Liang W, Luo S, He B, Li Q, Yeung SLA. Factors contributing to higher coronavirus disease 19 (COVID-19) risk or its severity: a living systematic review of Mendelian randomization study. PROSPERO 2021 CRD42021252079. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021252079 (23 June 2022, date last accessed).
116. Zhang X, Gill D, Li X, *et al.* Non-genetic biomarkers and risk of colorectal cancer: an umbrella review of meta-analyses of observational studies, meta-analyses of randomised controlled trials and Mendelian randomisation studies. PROSPERO 2019 CRD42019119833. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019119833 (23 June 2022, date last accessed).
117. Cheng TS, Day F, Lakshman R, Ong K. Association of puberty timing with type 2 diabetes: systematic review and meta-analysis. PROSPERO 2019 CRD42019124353. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019124353 (23 June 2022, date last accessed).
118. von Elm E, Altman DG, Egger M *et al.*; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453–57.