Open access • Posted Content • DOI:10.1101/2021.10.21.21265126

Tools for the assessment of quality and risk of bias in Mendelian randomization studies: a systematic review — Source link

Francesca Spiga, Mark Gibson, Sarah Dawson, G Davey Smith ...+2 more authors

Institutions: University of Bristol

Published on: 25 Oct 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Systematic review

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

Francesca Spiga,[1,2] Mark Gibson,[2,3] Sarah Dawson,[1] George Davey Smith,[1,2] Marcus R Munafò,[2,3] Julian PT Higgins[1,2]

[1] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[2] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[3] School of Psychological Science, University of Bristol, Bristol, UK

Corresponding author:
Francesca Spiga: f.spiga@bristol.ac.uk

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 risk of bias and/or quality of evidence assessment 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. 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.

Conclusions. We present an overview of tools and methods to assess risk of bias/quality of evidence in MR analysis. As none of these methods has been tested and validated for general use, we do not provide recommendations on their 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.

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, GI) 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 exact,[4] 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.[5]

MR was introduced as a way of strengthening causal inference regarding the kinds of modifiable exposures studied in conventional observational epidemiological studies. The key assumption here is that differences in an exposure induced by the GI will produce the same downstream effects on health outcomes as differences in the exposure produced by environmental influences (gene -environment equivalence assumption).[6]

As for instrumental variables analyses in general, the validity of an estimate from a MR analysis relies on the GI satisfying three core assumptions: (1) the GI must be associated with the exposure (*IV1-relevance*), (2) there are no unmeasured confounders of the GI-outcome association (*IV2-independence*) and (3) the GI-outcome association must be mediated entirely via the exposure (*IV3-exclusion restriction*). Additional assumptions, which are variety of the fourth IV assumption (*IV4*),[7] may be required for some inferences: *i)* the association of the GI and the exposure and the effect of the exposure on the outcome are the same for all participants in the sample (*homogeneity*; *ii)* the GI 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] *iiii)* the differences in an exposure induced by the GI will produce the same downstream effects on health outcomes as differences in the exposure produced by environmental influences (*gene -environment equivalence assumption*).[6] The validity of two-sample MR studies, in which different samples are used to estimate the GI-exposure and GI-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) and the genetic variants are harmonised (i.e. they are in the same direction in the two samples).[9]

Violation of any of the underlying assumptions may lead to spurious or biased estimates, as may other features of the study. 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,[10] bias due to horizontal pleiotropy[11]) and biases related to the population from which the data are collected (e.g., bias due to population stratification,[1, 12, 13] bias due to sample overlap in two-sample MR).[14] For example, 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.[12] Furthermore, using weak instruments in MR analysis can lead to estimates biased toward the confounded exposure-outcome association (in one-sample MR) or toward the null (in two-sample MR).[10]

Since prominent expositions of the use of MR in epidemiology from 2003 onwards,[1] the use of MR has increased 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. While numerous tools for risk-of-bias assessment in studies of interventions have been developed for both RCTs[15] and non-randomized studies of intervention,[16] 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. We undertook two related sub-reviews: a comprehensive and objective review of tools for the systematic assessment of the

conduct, evaluation and/or reporting of MR studies; and 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 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 1July 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 **appendices 1 and 2**.

Study selection

Search results were managed using Endnote and Excel. Titles and abstracts were screened by one review author (FS) using Rayyan software (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 (FS and MG) 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 finalised data extraction form was compiled. From each article, the following general information was extracted by one review author (FS): 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 two 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 bias were addressed, how bias were addressed, if a narrative description of MR-specific bias was reported, and what bias where 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 single item. We mapped items across tools to examine how similar bias were addressed by different 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 identified in the second sub-review, 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 (FS).

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 20 were retrieved for full-text screening. The inclusion criteria were met by 13 articles (reporting 14 tools) that are included in this review. Flow diagram of identification, screening and inclusion of articles is shown in **figure 1**. Of the 13 included articles, six were identified from searches of electronic databases of peer-reviewed articles and four from searches of preprints archives and Google Scholar, two from cited references, two from searches of systematic reviews (search 2) and one from searches of protocols of systematic reviews (search 3). A list of the included tools is reported in **supplementary table 1**.

Of the 14 articles tools included, eight tools were designed for single use in a specific systematic review (seven reviews and one protocol) and six 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, eight tools had a single purpose, of which four were aimed at the conduct of MR studies, three were aimed at the reporting of MR studies and one was aimed at evaluation of MR studies. The remaining six tools had two purposes: evaluation and reporting MR studies.

Details of the seven tools designed (or used) for evaluation of MR studies are reported in table 1. Of these, Burgess,[17] Davies,[18] Grau-Perez[19] and Treur[20] were structured by domains and items, whereas Kuźma,[21] LS Lee[22] and Mamluk[23] were structured by items only. The number of domains within the first four 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 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 evaluation tools by classifying each item into a bias/topic domain, and then we assigned each item to a specific bias/topic. We found that of the 121 items among all 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, (four items were designed to address both evaluation of risk of bias and other aspect of MR analysis); of the 81 items designed to evaluate MR studies, 77 addressed only one bias and four addressed multiple biases.

Details of the biases addressed by each tool are reported in table 2. Of the 81 items addressing bias, 32 related to the three core IV assumptions. Ten items in seven tools addressed bias related to the relevance assumption (IV1), eight items in six tools addressed bias related to the independence assumption (IV2) and 14 items in seven tools addressed bias related to the exclusion restriction assumption (IV3). In addition, 11 items in four tools addressed bias related to the selection of the genetic instrument and 14 items in six tools addressed bias related to the selection of the population(s) or sample(s); five items in four tools addressed bias related to sensitivity analysis, 19 items in three tools addressed bias related to measurement errors and misclassification, two items in one tool addressed bias due to missing data, four items in three tools addressed bias due to other type of confounding and two items in one tool addressed other source of bias. We provide details of the 44 items addressing other aspect of the MR analysis, including items addressing the reporting of MR analysis, in supplementary table 2. 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 dataset(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 evaluation 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 3**. The number of domains ranged from three to six in the reporting tools and from five to ten in the conducting tools; the number of items ranged from seven 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, Boef[24] contained items addressing linkage disequilibrium and canalization; Davey Smith[25] contained items addressing homogeneity and sample overlap (in two-sample MR); Lor[26] contained items addressing linkage disequilibrium and heteroscedasticity. Among the conducting tools, Burgess,[17] Grover[27] and Lawlor[28] contained items addressing the three IV core assumptions, and variant harmonization; in addition, Burgess[17] contained one items addressing the homogeneity assumptions and Grover[27] and Lawlor[28] contained items addressing sample overlap; Swerdlow[29] contained items addressing linkage disequilibrium and horizontal pleiotropy.

Systematic reviews of MR studies

Completed reviews

A total of 2036 record were identified from searches 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**. Of the 35 included reviews, 25 were systematic reviews and ten were umbrella reviews. Of the 35 included reviews, 29 addressed a clinical

question (i.e., included studies on the casual effect of an exposure *vs* an outcome), and six reviews addressed a methodological question (e.g., the status of reporting in MR studies); 17 reviews reported MR studies only, 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: six reviews conducted risk-of-bias assessments only, five reviews conducted quality of evidence assessments only and three 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 4.**

A structured risk-of-bias tool for was used in five reviews: four of these (Grau-Perez,[19] Kuzma,[21] Mamluk[23] and Treur[20]) used tools developed specifically for risk-of-bias assessment in MR studies that are included in the above sub-review of tools (see **supplementary rable 1** and **table 2**); the fifth, Cheng,[30] used the Newcastle Ottawa Scale (NOS) for cohort studies[31] which was not specifically developed for MR studies. Four further reviews conducted risk-of-bias assessments but did not use a structured tool: Markozannes[32] and X Zhang[33] assessed horizontal pleiotropy; Pearson-Stuttard[34] addressed the selection of the genetic instrument(s); and Riaz[35,36] conducted evaluation of the three core assumptions.

Of the eight reviews that conducted a quality of evidence assessment, Markozannes[32] and Pearson-Stuttard[34] used a structured method based on statistical significance of the effect estimate and X Zhang[33] 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, Bochud[37] based the assessment of quality of evidence on the strength of the genetic variant; Firth[38] based the assessment on the results of the statistical analysis, the use of sensitivity analysis and test for bidirectional effects; Kim[39] based the assessment on statistical power; Kohler[40] based the assessment on the proportion of variance in risk factors explained by genetic instruments used and Li[41] 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 3.** 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),[12] 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), six reviews addressed bias related to the selection of the population or sample (i.e., population heterogeneity and selection bias), eight reviews addressed bias due to canalization, and four 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), five reviews addressed reverse causation (or bidirectionality), three reviews addressed the inability to assess non-linear associations, two reviews addressed statistical analysis and lack of genetic instrument, respectively, and one 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, therefore 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.** Five of the 45 included protocols were of published systematic review that were included in our sub-review of systematic reviews above.[42-46] 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 one 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, Ibrahim[47] and Verdiesen[48] planned to use STROBE-MR[25] and other published literature, including the MR guidelines by Davies,[18] LS Lee[22] planned to use a self-developed questionnaire (also included in our synthesis of tools) based on published guidelines including Davies,[18] Grover,[27] and Burgess.[17] Markozannes[49] planned to use a self-developed tool based on the results of the main analysis and of the sensitivity analysis; Naassila[50-52] planned to use Q-GENIE; Shi[53,54] planned to use a modified version of a recently developed tool (no reference provided); Visontay[55,56] planned to use the tool developed by Mamluk[23] and Wong[57] planned to conduct risk of bias assessment based on the guidelines from Davies.[18] 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, [58,59] Jiang[60] and van Oort[61] referred to the MR methods protocol published by Grover,[27] and Julian[62] did not report any reference. Of the remaining four protocols, Saribaz[63] planned to develop a risk-of-bias assessment method at the time of conducting the review; M Lee[64] planned to perform a descriptive assessment of the MR methods and of the genetic variants used in included studies; Luo[65] planned to perform an assessment based on sensitivity analysis methods and different choices of genetic variants as instrumental variables; Treur[45] planned to perform an assessment based on sensitivity analysis methods, on the choice of genetic variants, on the presence of sample 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 Mamluk[44] 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,[42] Dack,[66] Fell,[67] Haan,[68,69] Lemus[70] and Suh[71] planned to use NOS,[31] and Baldwin,[72] Cara[73] and Gianfredi[74] planned to use a modified version of NOS; Elsakloul[75] planned to use STROBE,[76] Fan[77] 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[78,79] planned to use ROBINS-I,[16] Yan[80] planned to use the ROB-2[15] and the ROBINS-I[16] tools, Wang[81] planned to use the Cochrane risk-of-bias assessment tool (no details provided).

Discussion

Our systematic review of tools developed for the conduct, evaluation and/or reporting of MR studies identified 14 instruments. 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,[19-23] whereas only two were developed for general use.[17,18] Despite notable variability in the structure and content of the 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, and 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 tools addressed bias due to measurement errors and only one tool addressed bias due to other sources including missing data. While 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.

Consistent 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% the approach or methodology used was 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,[25] Q-GENIE,[82] a self-developed tool included in our synthesis of tools[22] and a tool developed within another systematic review.[23] One review protocol planned to use a recently developed tool, that, similarly to the tool developed by Mamluk,[23] 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 Davies[18] and Grover.[27]

Our review has strengths and limitation. 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, for each objective, specific search string developed with the assistance of an information specialist. However, as some of the tools we have identified were developed within other type 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, although we classified linkage disequilibrium as relevant to the choice of genetic variant, it can introduce both confounding[5] and horizontal pleiotropy[83] and therefore it can be considered as bias related to the IV2 or IV3 domains.

By summarising 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. Given that none of the tools identified by our searches appears to have been formally tested, we are not able to make a

recommendation on what tool(s) should be adopted to assess MR studies. However, 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, and what types of biases reviewers should consider when assessing the quality of the evidence reported by individual MR studies.

Funding

F.S. and M.M. were supported by a Cancer Research UK (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme). M.M. is part of the Medical Research Council Integrative Epidemiology Unit (MRC IEU) at the University of Bristol (MM_UU_00011/7). This work was supported by Cancer Research UK (18281/A29019).

References

1. 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.
2. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33**(1): 30-42.
3. 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**(12): e352.
4. 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**(2): 99-111.
5. 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**(8): 1133-63.
6. Davey Smith G. Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice? *Int J Epidemiol* 2012; **41**(1): 236-47.
7. 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**(3): 214-20.
8. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; **29**(4): 722-9.
9. 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**(6): 1717-26.
10. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011; **30**(11): 1312-23.
11. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**(R1): R89-98.

12. Brumpton B, Sanderson E, Heilbron K, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* 2020; **11**(1): 3519.
13. 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 Research* 2017; **2**(11).
14. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; **40**(7): 597-608.
15. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019; **366**: l4898.
16. 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.
17. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations [version 2; peer review: 2 approved]. *Wellcome Open Research* 2020; **4**(186).
18. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; **362**: k601.
19. Grau-Perez M, Agha G, Pang Y, Bermudez JD, Tellez-Plaza M. Mendelian Randomization and the Environmental Epigenetics of Health: a Systematic Review. *Current environmental health reports* 2019; **6**(1): 38-51.
20. 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: 1-32.
21. Kuźma E, Hannon E, Zhou A, et al. Which Risk Factors Causally Influence Dementia? A Systematic Review of Mendelian Randomization Studies. *Journal of Alzheimer's disease : JAD* 2018; **64**(1): 181-93.
22. 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. 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**(6): 1972-95.
24. Boef AGC, Dekkers OM, Le Cessie S. Mendelian randomization studies: A review of the approaches used and the quality of reporting. *International Journal of Epidemiology* 2015; **44**(2): 496-511.
25. 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.
26. Lor GCY, Risch HA, Fung WT, et al. Reporting and guidelines for mendelian randomization analysis: A systematic review of oncological studies. *Cancer Epidemiology* 2019; **62**: 101577.
27. Grover S, Del Greco M. F, 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. 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* 2019.

29. Swerdlow DI, Kuchenbaecker KB, Shah S, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *International Journal of Epidemiology* 2016; **45**(5): 1600-16.
30. 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**(1): e1003017.
31. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *http://wwwohrica/programs/clinical_epidemiology/oxfordasp* 2013.
32. 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**(1): 11-36.
33. Zhang X, Gill D, He Y, et al. Non-genetic biomarkers and colorectal cancer risk: Umbrella review and evidence triangulation. *Cancer medicine* 2020; **9**(13): 4823-35.
34. 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.
35. 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 network open* 2018; **1**(7): e183788.
36. 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. *Journal of the American College of Cardiology* 2018; **71**(11).
37. Bochud M, Rousson V. Usefulness of Mendelian Randomization in Observational Epidemiology. *International Journal of Environmental Research and Public Health* 2010; **7**(3): 711-28.
38. 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 : official journal of the World Psychiatric Association (WPA)* 2020; **19**(3): 360-80.
39. Kim MS, Kim WJ, Khera AV, Won H-H. Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis. *medRxiv* 2020: 2020.08.18.20176578.
40. 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. *Journal of Psychiatric Research* 2018; **103**: 189-207.
41. 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.
42. Cheng TS, Day F, Lakshman R, Ong K. Association of puberty timing with Type 2 diabetes: systematic review and meta-analysis. PROSPERO 04/02/2019.
43. Kim MS, Jun KW. Obesity as a risk factor for cardiovascular outcomes: umbrella review of the literature. PROSPERO 11/04/2020.
44. 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 12/01/2015.
45. 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 23/04/2019.

46. 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 16/12/2018.
47. Ibrahim M, Golledge J, Singh T, Thanigaimani S, Morris D. Systematic Review and Meta-Analysis of Mendelian Randomisation of Abdominal Aortic Aneurysms. PROSPERO 09/08/2020.
48. Verdiesen R, Schmidt M. Causal risk factors for breast cancer: a systematic review and meta-analysis of Mendelian randomization studies. PROSPERO 23/02/2021.
49. Markozannes G, Kanellopoulou A, Dimopoulou O, et al. Systematic review of Mendelian randomization studies on risk of cancer. OSF. March 31 https://doi.org/10.17605/OSF.IO/2RUCT.
50. Naassila M, Bouajila N, Aubin HJ, et al. Alcohol intake and cancers, neurological, cardiovascular and liver diseases: systematic review implementing Mendelian randomization studies PROSPERO 30/03/2021.
51. Naassila M, Naouras B, Aubin HJ, et al. Effects of alcohol intake on the risk of cardiovascular diseases: systematic review PROSPERO 30/04/2021.
52. Naassila M, Naouras B, Aubin HJ, et al. Effects of alcohol intake on the risk of neurological diseases: systematic review PROSPERO 27/04/2021.
53. 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 21/10/2020.
54. 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, October 26.osf.io/8rtph.
55. Visontay R, Sunderland M, Slade T, Wilson J, Mewton l. The relationship between level of alcohol consumption and health outcomes: A systematic review of observational studies employing approaches to improve causal inference. PROSPERO 13/05/2020.
56. 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*; **11**(3): e043985.
57. 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 29/04/2021.
58. Grover S, Del Greco F, König IM. Identification of causal risk factors for neurodegenerative diseases using a mendelian randomization approach. PROSPERO 16/03/2018.
59. 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. *Systematic Reviews*; **7**(1): 145.
60. Jiang M, Xu W, Liu G. A systematic review of Mendelian randomization studies in lung cancer. PROSPERO 16/05/2019.
61. 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 19/09/2019.
62. Julian T, Cooper-Knock J, Shaw P. Mendelian randomisation in neurodegenerative disease: a systematic review. PROSPERO 16/11/2020.
63. 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 04/12/2020.

64. Lee M, McGuinness LA. A systematic review of Mendelian randomization studies using adiposity as an exposure. PROSPERO 15/05/2018.
65. Luo S, Zhao J, Yang Q, Schooling M. Mendelian randomization studies compared to randomized controlled trials: a systematic review and meta-analysis. PROSPERO 04/08/2017.
66. 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 16/11/2020.
67. 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 25/11/2020.
68. 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, September 16. https://doi.org/10.17605/OSF.IO/D9WZK.
69. 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 24/04/2018.
70. 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 26/05/2021.
71. Suh JW, Sun S, Chadni MJ, et al. Risk factors for cardiovascular multimorbidity: a systematic review of prospective studies. PROSPERO 02/03/2021.
72. 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 20/05/2020.
73. 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 13/08/2020.
74. 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 19/04/2019.
75. Elsakloul F, Dawson J. The genetic determinants of serum uric acid and their association with cardiovascular diseases: a systematic review. PROSPERO 05/04/2016.
76. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; **370**(9596): 1453-7.
77. Fan H, Chen Y, Lee YL. Habitual Coffee Consumption and Lung Function Decline: Multi-country Meta-analyses and Mendelian Randomization Studies. PROSPERO 07/05/2020.
78. Karwatowska 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 13/02/2020
79. Karwatowska 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*; **10**(9): e038258.
80. 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 17/07/2020.
81. 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 06/11/2018.

82. 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.
83. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018; **27**(R2): R195-R208.
84. Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *International journal of epidemiology* 2015; **44**(2): 496-511.
85. 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.
86. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafo MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014; **13**: 99-106.
87. Abbasi A. Mendelian randomization studies of biomarkers and type 2 diabetes. *Endocrine Connections* 2015; **4**(4): 249-60.
88. Abbasi A, Sahlqvist AS, 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**(10): e0163721.
89. Belbasis L, Bellou V, Evangelou E, Tzoulaki I. Environmental factors and risk of multiple sclerosis: Findings from meta-analyses and Mendelian randomization studies. *Multiple Sclerosis Journal* 2020; **26**(4): 397-404.
90. Belbasis L, Kohler CA, Stefanis N, et al. Risk factors and peripheral biomarkers for schizophrenia spectrum disorders: an umbrella review of meta-analyses. *Acta Psychiatrica Scandinavica* 2018; **137**(2): 88-97.
91. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS ONE* 2018; **13**(3): e0194127.
92. 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**(2): e14399.
93. Carnegie R, Zheng J, Sallis HM, et al. Mendelian randomisation for nutritional psychiatry. *The Lancet Psychiatry* 2020; **7**(2): 208-16.
94. 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**(1): 130-42.
95. Frayling TM, Stoneman CE. Mendelian randomisation in type 2 diabetes and coronary artery disease. *Current Opinion in Genetics and Development* 2018; **50**: 111-20.
96. Hu Q, Hao P, Liu Q, et al. Mendelian randomization studies on atherosclerotic cardiovascular disease: evidence and limitations. *Sci China Life Sci* 2019; **62**(6): 758-70.
97. Kei A, Koutsouka F, Makri A, Elisaf M. Uric acid and cardiovascular risk: What genes can say. *International Journal of Clinical Practice* 2018; **72**(1): e13048.
98. Kuzma E, Hannon E, Zhou A, et al. A systematic review of mendelian randomization studies investigating causal associations between risk factors and dementia. *Alzheimer's and Dementia* 2017; **13**(7): P1180.
99. Mamluk L, Edwards HB, Savovic J, et al. Prenatal alcohol exposure and pregnancy and childhood outcomes: A systematic review of alternative analytical approaches. *The Lancet* 2016; **388**: 73.

100. 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. *International journal of epidemiology* 2019; **48**(5): 1425-34.
101. 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. *Psychopathology Review* 2016; **a4**(1): 4-25.
102. Robinson PC, Choi HK, Do R, Merriman TR. Insight into rheumatological cause and effect through the use of Mendelian randomization. *Nature Reviews Rheumatology* 2016; **12**(8): 486-96.
103. Sommer A, Twig G. The Impact of Childhood and Adolescent Obesity on Cardiovascular Risk in Adulthood: a Systematic Review. *Current diabetes reports* 2018; **18**(10): 91.
104. Vasta R, D'Ovidio F, Logroscino G, Chio A. The links between diabetes mellitus and amyotrophic lateral sclerosis. *Neurological Sciences* 2021; **42**(4): 1377-87.
105. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia* 2020; **63**(11): 2359-71.
106. Zhang Z, Thorne JL, Moore JB. Vitamin D and nonalcoholic fatty liver disease. *Current Opinion in Clinical Nutrition and Metabolic Care* 2019; **22**(6): 449-58.
107. 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 23/10/2020.
108. Desai R, Stott J. Risk factors and dementia: systematic review of Mendelian randomization studies PROSPERO 13/05/2021.
109. Gibson M, Spiga F, Richmond R, Munafo MR. A systematic review of the reporting quality of studies that use Mendelian randomisation in UK Biobank. OSF 2021, April 7. https://doi.org/10.17605/OSF.IO/HN7WD.
110. Kim MS, Lee I, Park S. Impact of obesity on gastroenterological diseases: umbrella review of systematic review and meta-analysis. PROSPERO 17/09/2020.
111. Kim MS, Park S. Impact of obesity on renal and genitourinary outcomes: an umbrella review of systematic reviews and meta-analyses. PROSPERO 23/03/2021.
112. 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 20/08/2020.
113. Maretzke F, Bechthold A, Egert S, et al. Role of vitamin D in preventing and treating selected extraskeletal diseases – an umbrella review. PROSPERO 09/07/2018.
114. Romo MV, Huang JV, Lun S, Yeung A, Zhao J, Schooling MC. The conduct and reporting of Mendelian randomization studies: a systematic review. PROSPERO 03/02/2018.
115. 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 08/10/2018.
116. 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 05/04/2018.

Table 1: Details of tools designed/used for assessing risk of bias/evaluating quality in MR 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[17]	To provide guidelines for performing MR investigations. To provide advice on which analyses to perform in a MR investigation.		9	22
Davies[18]	To provide explanations of core concepts and recent developments in MR methods.		6	19
Grau-Perez[19]	To conduct a systematic review of MR studies evaluating the causal role of environmentally responsive DNAm changes on the development of health states.	Boef et al.[84]	6	28
Kuźma[21]	To conduct a systematic review of MR studies investigating causal relationships between risk factors and global cognitive function or dementia.	Q-Genie[82]	-	11
Lee LS[22]	To perform an updated systematic review and meta-analysis of MR that will provide further insight into the causative factors of dementia.	Davies et al.[18] Grover et al.[59] Burgess et al.[17]	-	11
Mamluk[23]	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 et al.[85] Lawlor et al.[13] Taylor et al.[86]	-	5
Treur[20]	To review evidence from studies that applied MR to assess causal effects between poor mental health and substance use.	STROBE-MR[25]	5	25

Abbreviations: DNAm=DNA methylation; MR=Mendelian randomization.

Table 2: Details specific MR bias and limitation addressed by items or questions within each assessing tool

Bias (or topic) domain	Specific bias (or topic)	Burgess	Davies	Grau-Perez	Kuzma	Lee	Mamluk	Treur	Total items
IV1-Relevance	Choice of variants	Y							1
	Weak instrument bias		Y	Y	Y	Y	Y	Y	9
IV2-Independence	Choice of variants	Y							1
	Confounding		Y	Y	Y	Y[b]	Y[c]	Y	7
	Population stratification						Y		1
IV3-Exclusion restriction	Choice of variants	Y							1
	Horizontal pleiotropy	Y	Y	Y	Y	Y	Y	Y	13
Genetic instrument	Choice of variants	Y							3
	Construction of genetic score							Y	1
	Variants harmonization	Y	Y			Y		Y	4
	Linkage disequilibrium	Y	Y			Y			3
Population/sample	Samples overlap[a]	Y	Y			Y		Y	5
	Population heterogeneity[a]	Y	Y	Y		Y	Y	Y	6
	Choice of controls			Y					1
	Selection bias			Y					1
Sensitivity analysis	Evidence of robustness	Y	Y			Y		Y	5
Measurement error	Exposure measurement error/misclassification			Y	Y			Y	13
	Outcome measurement error/misclassification			Y	Y			Y	6
Missing data	-			Y					2
Other confounding	-			Y	Y		Y		4
Other sources of bias	-				Y				2

[a]In two-sample MR analysis. [b]Confounding of the genetic instrument-outcome association. [c]Confounding of the genetic instrument-exposure association and of the genetic instrument-outcome association. Abbreviations: IV1=instrumental variable assumption 1; IV2=instrumental variable assumption 2; IV3=instrumental variable assumption 3.

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[87]	Systematic review and MR analysis	MR studies of biomarkers and T2D	Yes	28/0	No	N/A	No	N/A	Yes
Abbasi[88]	Systematic review	Biomarkers and T2D	No	17/122	No	N/A	No	N/A	Yes
Belbasi[89]	Umbrella review	Risk factors and peripheral biomarkers for schizophrenia and other psychotic disorders	No	5/36	No	N/A	No	N/A	No
Belbasi[90]	Umbrella review	Risk factors of multiple sclerosis	No	6/9	No	N/A	No	N/A	Yes
Bellou[91]	Umbrella review	Environmental risk factors and biomarkers for T2D	No	22/86	No	N/A	No	N/A	Yes
Bergmans[92]	Systematic review	Comorbid depression and T2D	No	4/12	No	N/A	No	N/A	Yes

Bochud[37]	Literature review on MR methods, applications, and limitations	MR studies	Yes	38/0	No	N/A	Yes	Strength of genetic variant	Yes
Boef[a84]	Systematic review	Methodology used in MR analysis	Yes	179/0	No	N/A	No	N/A	Yes
Carnegie[93]	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[30]	Systematic review and meta-analysis	Puberty timing and T2D and/or impaired glucose tolerance	No	1/27	Yes, Yes	Newcastle-Ottawa Scale [31]	No	N/A	No
Diemer[94]	Systematic review	Prenatal environment and offspring outcomes	Yes	43/0	No	N/A	No	N/A	Yes
Firth[38]	Umbrella review	Modifiable health behaviors and major mental disorders	No	12/32	No	N/A	Yes	Statistical analysis results, use of sensitivity analysis and	Yes

								test for bidirectional effects	
Frayling[95]	Systematic review	MR studies of T2D, coronary artery disease and hypertension	Yes	16/0	No	N/A	No	N/A	Yes
Grau-Perez[a19]	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[96]	Systematic review	MR studies of atherosclerotic cardiovascular disease	Yes	58/0	No	N/A	No	N/A	Yes
Kei[97]	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[39]	Umbrella review of systematic reviews and meta-analyses	Adiposity and cardiovascular disease events or mortality	No	27/11	No	N/A	Yes	Statistical power	Yes
Kohler[40]	Umbrella review of meta-analysis and MR	Environmental risk factors for depression	No	8/70	No	N/A	Yes	Proportion of variance in risk factors	No

	studies							explained by genetic instruments	
Kuzma[a21,98]	Systematic review	MR studies of risk factors and global cognitive function or dementia	Yes	18/0	Yes, Yes	Modified Q-Genie [82]	No	N/A	Yes
Li[41]	Umbrella review of systematic reviews and meta-analyses	Serum uric acid level and multiple health outcomes	No	36/101	No	N/A	Yes	Statistical significance of the effect estimate and statistical power	Yes
Lor[a26]	Systematic review	MR analyses in oncological studies	Yes	77/0	No	N/A	No	N/A	Yes
Mamluk[a23,99]	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[32]	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[100]	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[34]	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[101]	Systematic review	MR studies of psychopathology-related outcomes	Yes	19/0	No	N/A	No	N/A	Yes
Riaz[35,36]	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[102]	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
Sommer[103]	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[a29]	Review on methods for selecting instruments	MR studies	Yes	231/0	No	N/A	No	N/A	Yes

	for MR analysis and Systematic review of MR studies								
Treur[a20]	Systematic review	MR studies of poor mental health and substance use	Yes	63/0	Yes, Yes	Self-developed tool	No	N/A	Yes
Vasta[104]	Systematic review	Diabetes mellitus and amyotrophic lateral sclerosis	No	1/35	No	N/A	No	N/A	No
Yuan[105]	Systematic review and MR analysis	MR studies of risk factors of T2D	Yes	40/0	No	N/A	No	N/A	No
Zhang X[106]	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[33]	Systematic review	Vitamin D and non-alcoholic fatty liver disease	No	1/12	No	N/A	No	N/A	No

[a]Included in the synthesis of tools for the assessing/evaluating MR studies. [b]Based on the location of the SNPs. Abbreviations: DNAm=DNA methylation; MR=Mendelian randomization; N/A=not applicable; SNP=single nucleotide polymorphism; T2D=type 2 diabetes.

Table 4: List of included protocols of systematic reviews reporting MR studies.

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[107]	Whole blood ionized magnesium in healthy adults	Systematic review	No	No	N/A
Baldwin[72]	The impact of childhood maltreatment on mental health	Systematic review and meta-analysis	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale[31]
Cara[73]	Safety of enteral nutrition formulations with dietary fibre	Systematic review	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale[31]
Cheng[b42]	Puberty timing and T2D	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale[31]
Dack[66]	Early life exposure to mercury, growth and neurodevelopment	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale[31]
Desai[108]	Risk factors for dementia	Systematic review	Yes	Yes/Yes	Q-Genie[82]
Elsakloul[75]	Serum uric acid and cardiovascular diseases	Systematic review	No	NS/Yes	Pre-specified bespoke tool based on STROBE[76]
Fan[77]	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[67]	Maternal smoking and orofacial clefts	Systematic review and meta-analysis	No	NS/Yes	Newcastle-Ottawa Scale[31]
Gianfredi[74]	Physical activity and depression	Systematic review	No	NS/Yes	Adapted version of the Newcastle-Ottawa Scale[31]
Gibson[109]	Reporting quality in MR studies using UK Biobank data	Systematic review of MR studies	Yes	No	N/A
Grover[58,59]	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 (Grover et al. 2017)[27]
Haan[68,69]	Alcohol, tobacco and caffeine consumption in pregnancy and externalising disorders in offspring	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale[31]
Ibrahim[47]	MR studies of abdominal aortic aneurysms	Systematic review of MR studies and meta-analysis	Yes	Yes/Yes	STROBE-MR[25] and other publications.
Jiang[60]	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 published method protocol (Grover et al. 2017)[27]. Assessment of the robustness and credibility of the data synthesis using sensitivity analysis.
Julian[62]	MR studies of neurodegenerative	Systematic review of MR	Yes	Yes/No	Assessment of risk of bias, evidence base for methodological strengths and weaknesses

	disease	studies			using the published literature
Karwatowska[78,79]	Risk factors for disruptive behaviours	Systematic review and meta-analysis	No	NS/Yes	Adapted version of the ROBINS-I checklist[16]
Kim[43]	Obesity and cardiovascular outcomes	Umbrella review	No	No	N/A
Kim[110]	Obesity and gastroenterological diseases	Umbrella review	No	No	N/A
Kim[b111]	Obesity and renal and genitourinary outcomes	Umbrella review	No	No	N/A
Lee LS[a22]	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[64]	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[70]	T2D and incidence of 17 types of cancer	Systematic review and meta-analysis	No	NS/Yes	Newcastle-Ottawa Scale[31]
Liu[112]	Risk factors for coronavirus disease 19 (COVID-19)	Umbrella review	No	No	N/A
Luo[65]	MR studies compared to 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

Mamluck[b44]	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[c113]	Role of vitamin D in preventing and treating selected extra-skeletal diseases	Umbrella review	No	No	N/A
Markozannes[49]	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[51]	Alcohol intake and risk of cardiovascular diseases	Systematic review	No	Yes/Yes	Q-Genie[82]
Naassila[52]	Alcohol intake and risk of neurological diseases	Systematic review	No	Yes/Yes	Q-Genie[82]
Naassila[50]	Alcohol intake and cancers, neurological, cardiovascular and liver diseases	Systematic review	Yes	Yes/Yes	Q-Genie[82]
Romo[114]	Conduct and reporting of MR studies	Systematic review of MR studies	Yes	No	N/A
Saribaz[63]	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[53,54]	Prenatal Alcohol	Umbrella review	No	Yes/Yes	Modified recently developed tool (reference

	Exposure and Offspring Health Outcomes				not provided)
Solmi[115]	Safety and efficacy of cannabinoids and cannabis in treating medical conditions	Umbrella review	No	No	N/A
Solmi[116]	Psychosis and non-communicable general medical conditions	Umbrella review	No	No	N/A
Suh[71]	Risk factors for cardiovascular multimorbidity	Systematic review	No	NS/Yes	Newcastle-Ottawa Scale[31]
Treur[b45]	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[61]	Alcohol consumption and its causal relationship with mortality, cardio-metabolic 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 (Grover et al. 2017)[27] with focus on MR design, the quality of the genetic instrument, and the validation of the MR assumptions
Verdiesen[48]	Causal risk factors for breast cancer	Systematic review of MR studies and meta-analysis	Yes	Yes/Yes	STROBE-MR[25] and a published checklist (Davies et. Al, 2018)[18]
Visontay[55,56]	Alcohol consumption and	Systematic	No	Yes/Yes	Recently developed risk of bias tools specific

	health outcomes	review			to MR studies, natural experiments, and other genetic-based methods (Mamluk et al., 2021)[23]
Wang[81]	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
Wong[57]	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 (Davies et. Al, 2018)[18]
Yan[80]	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 randomised controlled trials)[15] and the ROBINS-I[16]
Zhang[b46]	Non-genetic biomarkers and risk of colorectal cancer	Umbrella review	No	No	N/A

[a]Included in the synthesis of tools for assessing, conducting and reporting MR studies. [b]Protocols of published systematic reviews included in this article. [c]Protocol of published systematic review not included in this article. Abbreviations: MR=Mendelian randomization; N/A=not applicable; 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.

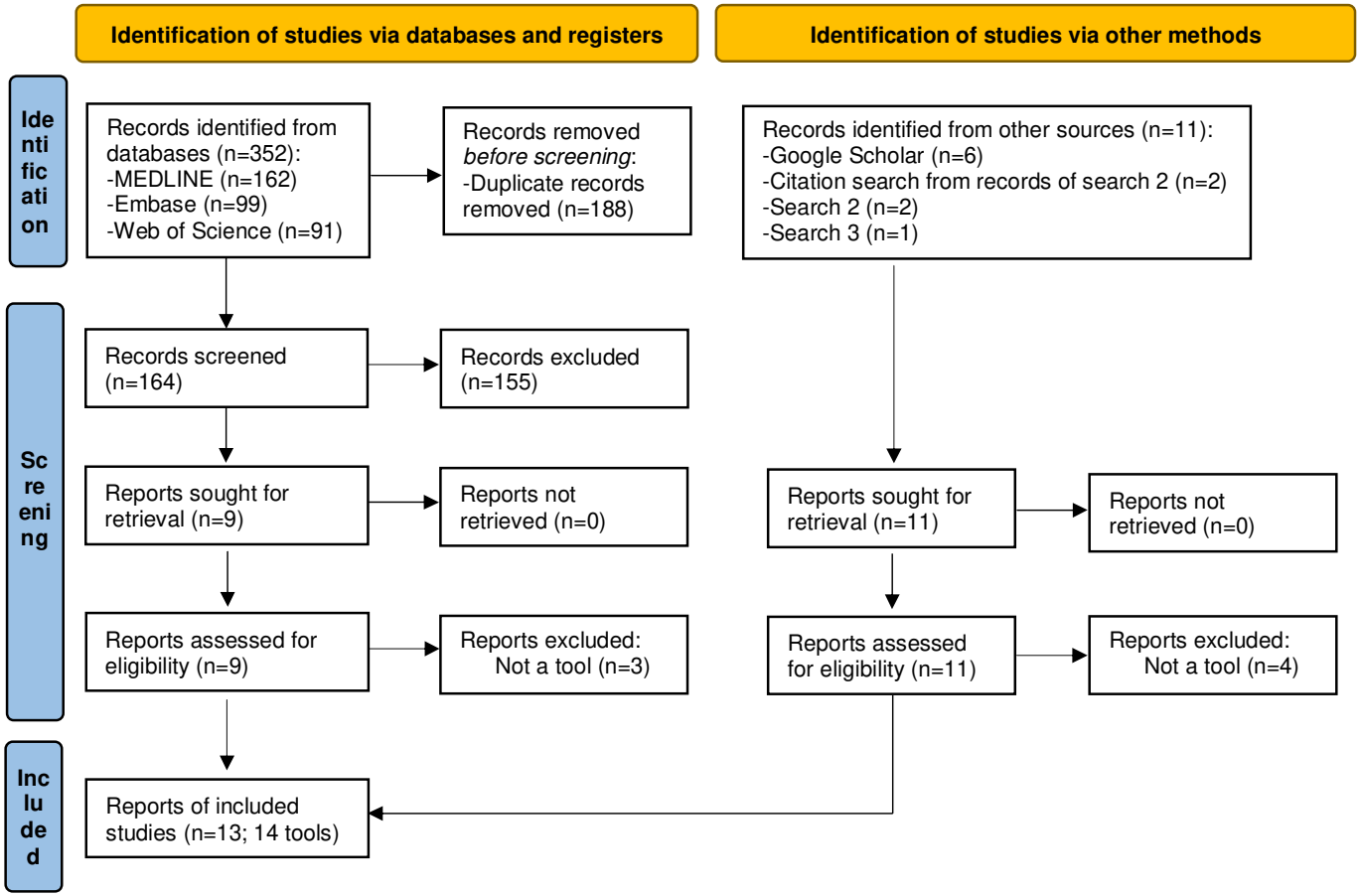

Figure 1: Flow diagram of identification, screening and inclusion of articles containing tools for assessing, conducting and/or reporting Mendelian randomization studies.

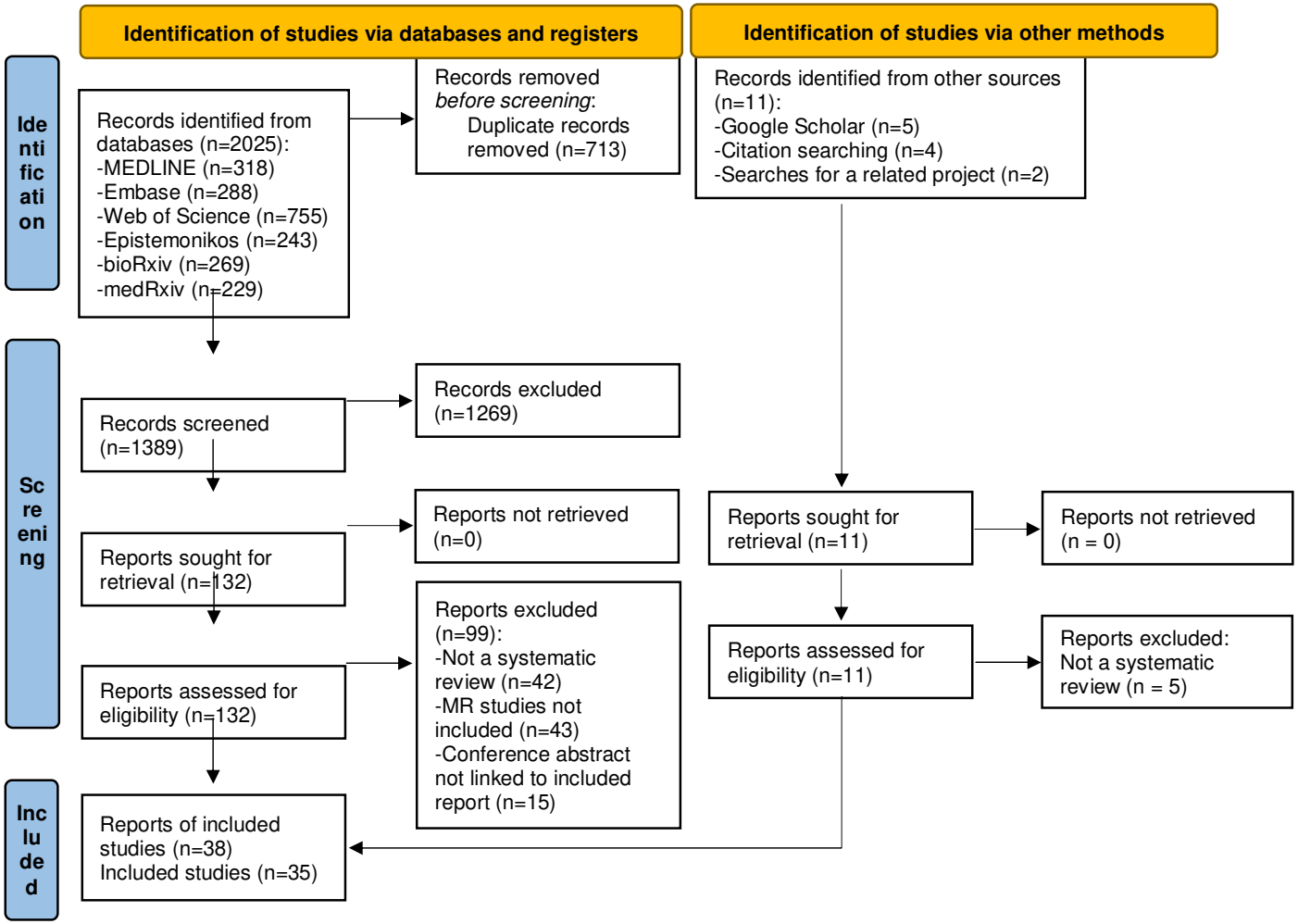

Figure 2: Flow diagram of identification, screening and inclusion of articles containing systematic reviews (and meta-analysis) of Mendelian randomization studies

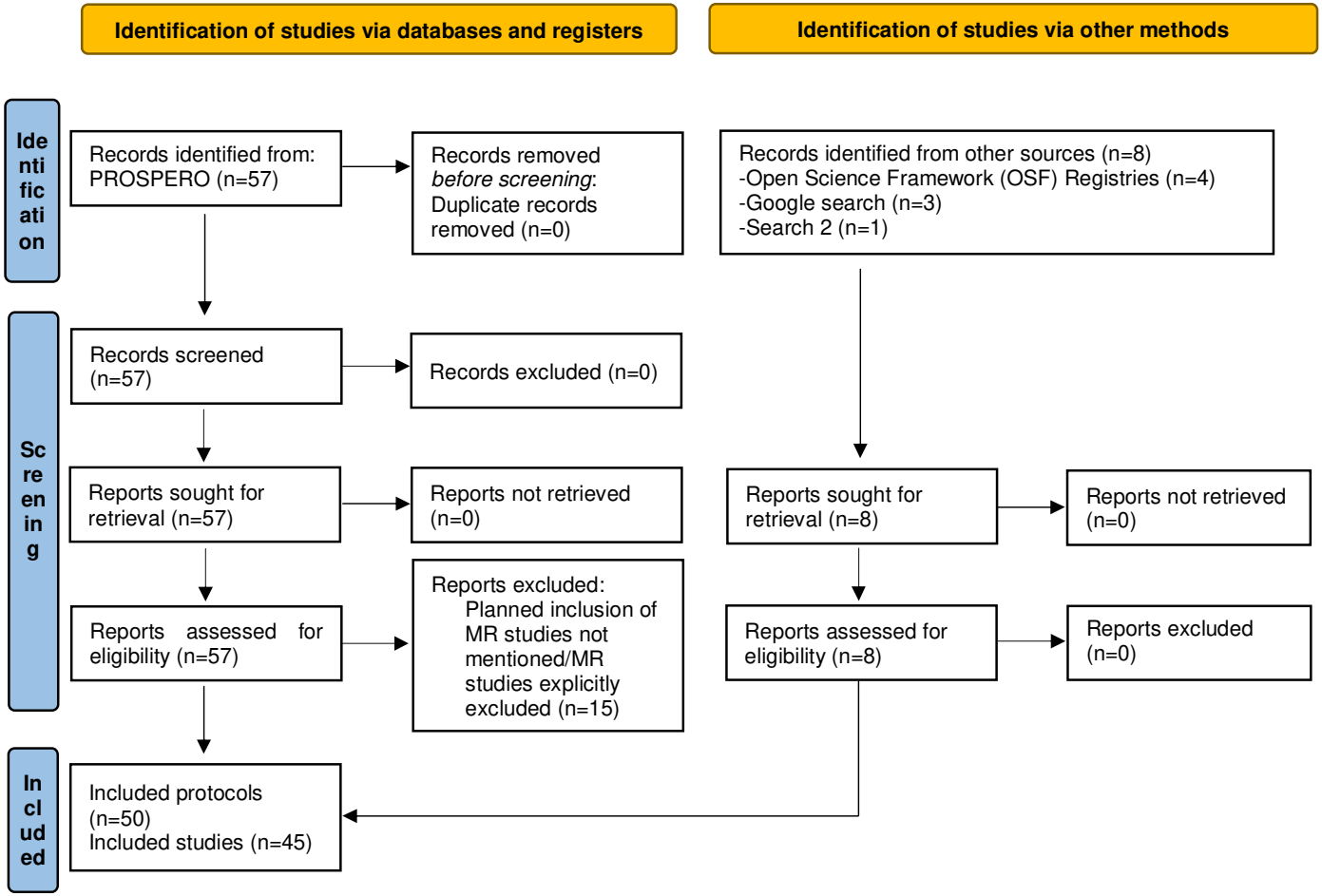

Figure 3: Flow diagram of identification, screening and inclusion of protocols of systematic reviews (and meta-analysis) planning to include Mendelian randomization studies.