BMJ Open

Systematic survey of the causal language use in systematic reviews of observational studies: a study protocol

Mi Ah Han ,[1] Gordon Guyatt[2]

To cite: Han MA, Guyatt G. Systematic survey of the causal language use in systematic reviews of observational studies: a study protocol. *BMJ Open* 2020;**10**:e038571. doi:10.1136/bmjopen-2020-038571

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-038571).

Received 16 March 2020
Revised 24 June 2020
Accepted 06 July 2020

[1]Department of Preventive Medicine, College of Medicine, Chosun University, Gwangju, The Republic of Korea
[2]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

Correspondence to
Dr Mi Ah Han;
mahan@chosun.ac.kr

ABSTRACT

Introduction Sometimes, observational studies may provide important evidence that allow inferences of causality between exposure and outcome (although on most occasions only low certainty evidence). Authors, frequently and perhaps usually at the behest of the journals to which they are submitting, avoid using causal language when addressing evidence from observational studies. This is true even when the issue of interest is the causal effect of an intervention or exposure. Clarity of thinking and appropriateness of inferences may be enhanced through the use of language that reflects the issue under consideration. The objectives of this study are to systematically evaluate the extent and nature of causal language use in systematic reviews of observational studies and to relate that to the actual intent of the investigation.

Methods and analysis We will conduct a systematic survey of systematic reviews of observational studies addressing modifiable exposures and their possible impact on patient-important outcomes. We will randomly select 200 reviews published in 2019, stratified in a 1:1 ratio by use and non-use of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE). Teams of two reviewers will independently assess study eligibility and extract data using a standardised data extraction forms, with resolution of disagreement by discussion and, if necessary, by third party adjudication. Through examining the inferences, they make in their papers' discussion, we will evaluate whether the authors' intent was to address causation or association. We will summarise the use of causal language in the study title, abstract, study question and results using descriptive statistics. Finally, we will assess whether the language used is consistent with the intention of the authors. We will determine whether results in reviews that did or did not use GRADE differ.

Ethics and dissemination Ethics approval for this study is not required. We will disseminate the results through publication in a peer-reviewed journals.

Registration Open Science Framework (osf.io/vh8yx).

Strengths and limitations of this study

- Our systematic survey will be the first to evaluate current practice of published systematic reviews of observational studies of causal language use.
- We will use robust methodology including a comprehensive sample of recent eligible reviews, explicit eligibility criteria, duplicate and independent eligibility screening and data abstraction, pilot testing of forms and, to ensure consistency and reproducibility, detailed instructions for making judgement regarding causal language use.
- Several reviewers will participate in this review and will make subjective judgements at each step of the process. Judgements regarding whether the authors did or did not use causal or association language, and particularly whether their intent was or was not to make causal inferences, may be challenging. We will provide detailed instructions and conduct piloting and calibration exercises, to minimise disagreement and maximise accuracy.

INTRODUCTION

In general, well-designed randomised controlled trials (RCTs) provide the best evidence for the causal relation between an exposure and outcome. In many instances, however, often because RCTs are not ethical or feasible, the best available evidence regarding causation comes from observational studies.

The true interest of investigators summarising the results of observational studies may be either in association or causation. For exposures that are not modifiable, such as age, sex and family history of disease, the issue is understanding prognostic power through the presence and magnitude of association.

For modifiable risk factors, however, the interest is typically in causation. For instance, because of the possibility of modifying diet, investigators may be interested in the relation between dietary factors and patient-important health outcomes. With respect to other volitional behaviours such as breast feeding, the interest is modifying behaviours with a view to improving patient or childhood outcomes. With respect to potentially toxic exposures such as smoking or radiation, investigators have in mind avoiding the exposure. Were associations non-causal, there would be no

point in modifying exposures because the modification would have no impact on outcomes.

Despite investigators conscious and clear—or sometimes less conscious and clear—focus on establishing causation with a view to intervening to improve outcomes, many journals ignore the underlying study question and request that authors avoid causal language based on the study design. For example, *JAMA*, a prestigious high-impact journal, requests that authors restrict causal language to RCTs and all other study designs, including systematic reviews of RCTs, should be described in terms of association or correlation.[1]

On occasion, observational studies provide compelling evidence of causation that mandates change in behaviour. For instance, we have been sufficiently convinced, largely because of the magnitude of association, to intervene to reduce harm from smoking.[2 3] Similarly, observational studies have provided evidence of causation sufficiently compelling to mandate legislation requiring use of seat belts in motor vehicles[4] and helmets for those riding bicycles.[5] In other instances, for example, red meat consumption and cardiovascular and cancer risk, investigators claim causation on the basis of much lower certainty evidence, and nevertheless take strong public health stances based on their beliefs.[6]

In all these situations, failure to use language that reflects the purpose of the studies, and the inference investigators and consumers of research draw, can only confuse and obscure the discussion.[7] For instance, authors who follow journal requirement and restrict themselves to non-causal (association) language regarding red meat consumption and health outcomes and then recommend reducing red meat consumption to improve outcomes that are manifesting an internal contradiction.

When the issue is truly association, language should reflect that objective, when the objective is establishing or refuting causation, the language should reflect that issue. Guidance from the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group, widely used for assessing the certainty of evidence from both RCTs and observational studies, vividly clarifies the issue. Using the GRADE approach, when the issue is establishing association with a view to clarifying prognosis (ie, risk factors such as age, sex and disease stage) bodies of evidence from observational studies begin as high certainty evidence.[8] When the issue is causation, observational studies begin as low certainty evidence.[9]

For example, a systematic review provided evidence of a link between nocturia and mortality. The authors addressed the issue of association (the presumption being that disease states that cause nocturia, and not the nocturia itself, is responsible for the association, and interventions directed at nocturia will not influence mortality) rating the evidence as moderate certainty. As part of the same presentation, they addressed the issue of causation (nocturia itself is responsible for mortality and reducing nocturia could favourably impact on mortality) rating the evidence as very low certainty.[10] The same was true for the outcome of falls and fractures.[11]

The objectives of this study are to evaluate the use of causal language in systematic reviews that focus on observational studies of modifiable exposures and patient-important outcomes and to relate that language to the actual objectives of the investigation (only establishing association or making causal inferences). Because we hypothesise that systematic reviews are more likely to use language appropriately according to the study question whether they use the GRADE, we will compare the results according to the use and non-use of the GRADE.

METHODS AND ANALYSIS

We will conduct a systematic survey of published systematic reviews of observational studies and will use standard methodology for conducting such surveys. Online supplementary appendix A presents the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols checklist.[12] We will document and publish important amendments to the protocol with the results of this study.

Eligibility criteria

We will include systematic reviews meeting all following criteria:

1. Include only observational studies such as cohort, case–control, cross-sectional and case–cohort studies.
2. Examine the association between one or more modifiable exposures and patient-important outcomes.
3. Published in English during 2019.

Modifiable exposures are those that are amenable to change through conscious action. Common examples include health behaviours (eg, smoking, alcohol consumption, physical activity, diet), preventive health service uptake (eg, cancer screening, vaccination), biological status (eg, blood pressure, blood lipids, diabetes, symptoms, function), therapeutic clinical intervention (eg, behaviour change facilitation, drug therapy, surgical therapy) and environmental exposure (eg, herbicide).

Non-modifiable exposures are those that people cannot change or change is very difficult or unusual such as age, gender, ethnicity, family history of disease and genetic characteristics (eg, mutation or expression of gene).

Patient-important health outcomes include mortality (eg, all-cause mortality, disease-specific mortality), morbidity (eg, cardiovascular disease, cancer, hospitalisation), quality of life (eg, overall, disease-specific quality of life), function and symptoms (eg, dyspnoea, pain).

We will exclude reviews including evidence from both RCTs and observational studies because language use would be affected by the use of both study designs. In particular, when authors include RCTs, it is much more likely that they will use causal language, and this will be a result of the inclusion of RCTs in the review; even though a review planned to include RCTs, but there are no eligible RCTs and only observational studies are available, the review will be not eligible.

We will exclude systematic reviews including only cross-sectional studies. However, if a review includes other observational study designs as well as cross-sectional studies, the review will be eligible.

We will also exclude systematic reviews addressing diagnostic accuracy (performance) and ecological studies. Narrative reviews, umbrella reviews, network meta-analysis, commentaries, letters and protocols not presenting original data will not be eligible.

Literature search

We will search EMBASE, MEDLINE and Epistemonikos to identify potentially eligible systematic reviews. Since we are interested in current status of how authors use the causal language of their systematic review, we will include the most recent studies. Therefore, we will limit the search in 2019 and, if we find more eligible reviews than required, we will randomly select reviews from among those eligible (see the Selection of eligible reviews section). If we do not reach the final sample size (see the Sample size section), we will include reviews published in 2018, or if necessary 2017 or further. Online supplementary appendix B presents the search strategy.

Review process

Teams of two reviewers will screen reviews for eligibility and extract data independently and in duplicate. Reviewers will resolve disagreement through discussion or, if necessary, through consultation with a third reviewer. To ensure the validity and consistency of the review process, we will conduct calibration exercises for each process until reviewers reach a high level of agreement. We will develop and pilot test standardised forms for eligibility screening and data extraction and provide reviewers with corresponding detailed written instructions.

When a review reports more than one eligible paired exposure and outcome of interest comparison, we will select the result reported first in the abstract assuming that reflects the authors' primary interest.

Selection of eligible reviews

In the title and abstract screening, reviewers will judge if the study may be a systematic review of observational studies evaluating the association between modifiable exposure and patient-important outcomes. If either reviewer thinks the study may meet eligibility criteria, we will obtain full texts. We will then judge eligibility of the full texts.

For all studies that meet eligibility on full-text screening, we will determine whether the authors do or do not rate the certainty of evidence using GRADE and will then treat use or non-use of GRADE as a stratification variable. We will determine the number of eligible articles in each stratum. If we identify more than 100 eligible studies published in 2019 in both strata, we will randomly select studies to meet our sample size requirement. If, in the GRADE stratum, we do not meet the sample size in 2019 studies (ie, we identify fewer than 100 studies), we will randomly select the number of non-GRADE reviews as GRADE reviews published in 2019 and will repeat the process searching for more eligible studies in 2018 and if necessary 2017 or further. This process will allow the same number of reviews with or without the GRADE for each publication year.

Data abstraction

We will collect study characteristics from each eligible systematic review including country of corresponding author, journal, protocol registration, study design of included observational studies (cohort, case–control, cross-sectional, case–cohort), number of included observational studies, number of included participants, modifiable exposure, wording for exposure (eg, intervention, exposure), primary outcome investigated, point estimate and confidence or certainty interval, source of funding and conflict of interest.

We will review the discussion for statements that convey the authors' intent regarding causation and, on this basis, make a decision regarding authors' intent. We will then examine language use in each of the abstract, introduction, methods and results and judge whether the language in each section is consistent with the intent. For papers that use GRADE, we will assess whether the GRADE use is consistent with the intent. For consistency and transparency, we will develop detailed guidance regarding classification of language as well as classification of intent. We will pilot test a draft data abstraction form with 10 eligible studies (5 of non-GRADE use and 5 of GRADE use). The detailed explanations are as follows.

The intent of the authors

We will evaluate the intent of the authors in the Discussion section. We will conclude that the authors' intent is to address causation if they reflect on the merits of use or non-use of an intervention to modify the exposure with the intent of modifying the outcome (table 1). Despite an actual intent to address causation, authors might avoid clear causal language due to journals' demands or their own habits. The Discussion section is most likely to include language that conveys causal assessment intent

Table 1 Criteria for judging the authors' intent on causation

Section	Criteria	Response
Discussion in main text	Did the authors reflect on the merits of use or non-use of an intervention to modify the exposure with the intent of modifying the outcome? (eg, recommendation regarding the merits of intervention or behaviour change in changing the outcome; comment on certainty regarding the merits of behaviour change or intervention in changing the outcome)	Yes or no

Table 2 Criteria for judging causal language use

Section	Criteria	Response
Title	Did the authors use clearly causal language reflecting direct influence of exposure or intervention on outcome? (eg, impact, effect, influence, determinant)	Yes or no
Objective in abstract	Did the authors mention that the aim of the review is investigating causality? (eg, direct influence of exposure on outcome)	Yes or no
Results in abstract	Did the authors describe the direct influence of exposure or intervention on outcome? (eg, result in, showed outcome increase, have little or no effect)	Yes or no
Conclusion in abstract	Did the authors conclude the direct influence of exposure or intervention on outcome? (eg, effect, impact; recommendation of intervention or behaviour change; comment on certainty regarding behaviour change or intervention)	Yes or no
Introduction in main text	Did the authors mention that the aim of the review is investigating causality? (eg, direct influence of exposure on outcome)	Yes or no
Results in main text	Did the authors describe the direct influence of exposure or intervention on outcome? (eg, result in, showed outcome increase, have little or no effect)	Yes or no

through recommending that modification of exposure be undertaken with a view to improving outcomes. This is why we will focus on the Discussion section to establish authors' intent. If there is no statement regarding the possible impact of modification of the exposure, we will conclude that the intent is inferences only regarding association.

Causal language use

We will divide the language into the language of causation versus the language of association. Causal language is used to indicate situations in which the exposure directly influences the outcome. The language of association refers only to the link between exposure and outcome and does not imply causation. The language of association implies that the outcome may not change even if the exposure changes.

We will assess whether the authors use causal language in six sections: title, objective in the abstract, results in the abstract, conclusion in the abstract, study question in the introduction of main text and results in main text (table 2).

GRADE use

For systematic reviews using GRADE to assess the certainty of the evidence, we will evaluate whether the authors start the certainty of evidence at high or low. If the authors have the intent to address causation and start the certainty starts at low (or rate down two levels for risk of bias if using ROBINS-I (Risk Of Bias in Non-randomized Studies - of Interventions)), we will judge their intent consistent with the GRADE guidance. If the authors' intent is addressing association and the certainty starts at high, we will judge their intent consistent with the GRADE use. In each case, if the opposite, we will judge their practice inconsistent with GRADE guidance.

Sample size

Since we select reviews in stratum by the GRADE use, we calculate sample size using two independent samples' test for proportion comparison of the causal language use. We assume 50% of reviews use causal language in the non-GRADE stratum and 70% of reviews use causal language in the GRADE use stratum. With an α of 0.05 and β of 0.2, 100 reviews per each stratum arm will power our study at over 80%.[13]

Analysis

We will conduct a descriptive analysis of all variables. We will provide a summary of the study characteristics including the number of included observational studies, number of participants, type of exposure and outcomes.

We will calculate the proportion of systematic reviews with intent on causation and will calculate the proportion of systematic review using causal language in the study title, abstract, study question and results according to the authors' intent. We will compare them according to the use and non-use of GRADE with the χ^2 tests.

We will conduct two multivariable logistic regression analyses to examine the association between prespecified study characteristics and, first, causal intent for all reviews included and, second, causal language use in each section of the article for reviews with causal intent.

We list and rank our prespecified study characteristics for the regression analyses by importance as follows: GRADE use (yes or no), journal (high-impact vs other journals), type of exposure (therapeutic clinical interventions vs other exposure), statistical significance of the main effects (statistically significant or not), study design of primary studies (cohort vs other design), number of participants included (continuous variable) and source of funding (partially or completely funded by private for-profit organisation vs others). If there are sufficient

events (causal intent and causal language use), we will include them all. Otherwise, we will include as many as possible according to the 10 events-per-category rule which requires 10 events per category.

We hypothesise that reviews are more likely to use causal language if they use GRADE, publish in higher impact journals (*Journal of American Medical Association*, *New England Journal of Medicine*, *The Lancet*, *British Medical Journal*, *Annals of Internal Medicine and Public Library of Science Medicine*), evaluate therapeutic clinical interventions, report statistically significant results, focus on cohort studies, have larger sample size and receive funding from for profit organisation.

Patient and public involvement

No patient involved.

DISCUSSION

Main objectives of our study

Our review will systematically evaluate the current status of causal language use in systematic reviews of observational studies. This review is important because the choice of language is a powerful means of conveying the authors' intentions whether their goal is limited to prediction/association versus their goal is to make causal inferences.

IMPLICATIONS

This protocol describes the methodology and details of a planned systematic survey addressing the causal language use in published systematic reviews of observational studies. The findings of this study will inform the systematic review community regarding the current practice of causal language use, will highlight the limitations of current practice and will provide an opportunity for suggesting improved guidelines for appropriate selection of language for future reviews. Our results will draw the attention of primary researchers as well as systematic review authors, guideline developers and journal editors.

Ethics and dissemination

Ethics approval is not required because we will only use published reviews. We will disseminate theresults of this review through the publication in a peer-reviewed journal.

Contributors MAH conceived the study, designed the study, drafted and critically reviewed manuscript and finalised the protocol. GG developed the study design, provided guidance to the study conceptualisation and protocol development, critically reviewed the manuscript and finalised the protocol.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Mi Ah Han http://orcid.org/0000-0003-1213-6952

REFERENCES

1 Instructions for authors: JAMA network, 2019. Available: https://jamanetwork.com/journals/jama/pages/instructions-for-authors [Accessed 22 Nov 2019].
2 Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung Cancer* 2004;45:S3–9.
3 Lucas R, Rodney Harris R. On the Nature of Evidence and 'Proving' Causality: Smoking and Lung Cancer vs. Sun Exposure, Vitamin D and Multiple Sclerosis. *Int J Environ Res Public Health* 2018;15:1–3.
4 Fouda Mbarga N, Abubakari A-R, Aminde LN, *et al*. Seatbelt use and risk of major injuries sustained by vehicle occupants during motor-vehicle crashes: a systematic review and meta-analysis of cohort studies. *BMC Public Health* 2018;18:1413.
5 Olivier J, Creighton P. Bicycle injuries and helmet use: a systematic review and meta-analysis. *Int J Epidemiol* 2017;46:278–92.
6 Johnston BC, Zeraatkar D, Han MA, *et al*. Unprocessed red meat and processed meat consumption: dietary guideline recommendations from the nutritional recommendations (NutriRECS) Consortium. *Ann Intern Med* 2019;171:756.
7 Hernán MA. The C-Word: scientific Euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018;108:616–9.
8 Iorio A, Spencer FA, Falavigna M, *et al*. Use of grade for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870.
9 Guyatt G, Oxman AD, Akl EA, *et al*. Grade guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
10 Pesonen JS, Cartwright R, Vernooij RWM, *et al*. The impact of nocturia on mortality: a systematic review and meta-analysis. *J Urol* 2020;203:ju0000000000000463.
11 Pesonen JS, Vernooij RWM, Cartwright R, *et al*. The impact of nocturia on falls and fractures: a systematic review and meta-analysis. *J Urol* 2020;203:ju0000000000000459.
12 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
13 Inference for proportions: comparing two independent samples. Available: https://www.stat.ubc.ca/~rollin/stats/ssize/b2.html [Accessed 12 Jun 2020].