Systematic reviews to evaluate causation: an overview of methods and application

Khalid S Khan,[1,2] Elizabeth Ball,[1,2] Caroline E Fox,[3,4] Catherine Meads[2]

[1]Department of Obstetrics and Gynaecology, Royal London Hospital, London, UK
[2]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK
[3]Birmingham Women's Hospital NHS Foundation Trust, Birmingham, UK
[4]Reproduction, Genes and Development, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

Correspondence to:
Khalid S Khan
Centre for Primary Care and Public Health, Queen Mary University of London, Yvonne Carter Building, 58 Turner St, Whitechapel, London, E1 2AB, UK;k.s.khan@qmul.ac.uk

Abstract

Currently systematic reviews focus on diagnosis or effectiveness of treatment. It is the understanding of disease aetiology that underpins medical education, practice and research. Whether an association meets causal criteria is usually assessed qualitatively. However, this can also be examined through evidence synthesis and systematic reviews to evaluate disease causation and mechanisms are much needed. It is important in such a review to specify the questions to be addressed regarding causal criteria such as strength, consistency, temporality, specificity, biological gradient, plausibility and experimental evidence. The next step is to conduct a thorough literature search to identify the relevant studies and to assess them for their quality, particularly in relation to the risk of bias, ascertainment of exposures and ascertainment of outcomes. Data synthesis can then examine if the observed associations in collated studies are consistent, strong and temporal using techniques such as meta-analysis, testing for heterogeneity and meta-regression. Biological plausibility and coherence with existing theories can also be examined systematically through an assessment of the basic scientific literature. Experimental evidence might also be collated and synthesised to determine if removal of a causal agent alters the outcome. Through these steps a systematic review can help to establish whether an association is causal or not.

Introduction

Good-quality systematic reviews inform evidence-based decision making, but they usually focus on diagnosis or effectiveness of treatment rather than the aetiology of a condition. The cause of many medical conditions is multifactorial and can be hard to establish[1 2] but it is the understanding of disease causation that underpins medical education, practice and research. Thus, systematic reviews to evaluate aetiological evidence are much needed. This article outlines methods employed to review causation, collating evidence on the relationship between exposures and putative clinical outcomes.

Assessing causation using systematic reviews

Systematic reviews are robust pieces of research based upon a clearly formulated question, from which it is possible to identify relevant studies, appraise their quality and summarise the evidence using scientific and replicable methods. It is the use of an explicit and systematic approach that differentiates such reviews from anecdotal evidence, expert commentaries and narrative reviews. A causation review requires specific steps to transparently investigate causal criteria (table 1).

Originally causal criteria such as strength, consistence and temporality of association (this is not a comprehensive list) were derived from studies in the fields of microbiology[3] and epidemiology[2] but have now gained wider acceptance within clinical medical research.[4–6] Causation reviews are difficult to identify because there are no specific medical subject headings (MESH terms) used for indexing within the searchable databases. However, there are MESH terms for aetiology (and causality) which can be combined with search filters for systematic reviews.[7]

In addition, evidence for causation of health disorders may be concealed within the results sections or tables of a published review and not necessarily labelled as such. For example, when applying the experimental evidence criterion (please refer table 1) the presumed causal agent may be removed within a randomised controlled trial. If the condition under investigation is absent after treatment compared with the control group where it is present, a causal relationship may be inferred. For example, the removal of visible areas of endometriosis and restoration of anatomy by division of adhesions in order to treat pelvic pain associated with endometriosis, which has been investigated in a recent Cochrane review.[8]

Setting up a causation systematic review

At the outset, any hypothesis concerning strength, consistency and temporality should be specified. The mechanisms behind the development of the condition under investigation should be explored and the disease process must be outlined including aetiological factors, pathophysiology and clinical manifestations.[9] Studying aetio-pathogenesis in this way involves evaluation of the proposed biological pathway in relevant animal, laboratory or human studies.[10] Multiple sources should be searched and the search terms should identify literature on the events in the biological pathway, including unpublished studies to avoid the considerable risk of publication bias in basic science literature.[9–11] Standard systematic review techniques need to be used to ensure rigour in the final conclusions regarding causality.[12 13]

Assessing study design and quality

Epidemiological principles relevant to studying causation without bias include prospective design; measurement of causal agent, correct temporality of the association and control for confounding, although it is acknowledged that this list is not exhaustive. In observational studies, adjustment using multivariable modelling should be suggested as a way of controlling for known confounding variables. There are no validated quality assessment checklists or scores for causality systematic reviews yet so any quality assessment will need to acknowledge debate on what constitutes good quality in this area. This is similar to another new area for systematic reviews such as diagnostic yield where quality assessment checklists are evolving.[14]

Table 1 Common criteria for causation and their assessment through systematic reviews

Causal criteria	Evaluative systematic review methods
Strength of association	Comparing the relationship between presumed aetiological factor and condition under investigation, generating measures of association for individual studies, for example, ORs, and pooling results in meta-analysis if appropriate,
Consistency	Consistency of individual results across different studies (settings, tests for condition under investigation, study designs), examined graphically by L'Abbe and Forrest plots and statistically by χ^2 test and I^2 statistic.
Temporality	What comes first – condition under investigation or presumed aetiological factor? Information needs to be sought from studies with observations prior to exposure to the presumed aetiological factor.
Specificity	Often there are many purported aetiological causes for the condition under investigation. Attempts should be made to study the contribution made by the presumed aetiological factor over and above that made by any other pathologies using subgroup and meta-regression analyses.
Biological gradient	Does an increase in the presumed aetiological factor worsen the condition under investigation? This question can be studied using subgroup and meta-regression analyses for example, using test result as an explanatory variable for symptom severity.
Plausibility, coherence and analogy	Is the causation of the condition under investigation by the presumed aetiological factor biologically plausible? Does it conflict with generally known facts of the natural history of the condition? Are there analogous causal relationships between the aetiological factor and the condition? Answers may lie in the discussion sections of selected articles, which likely include reference to basic scientific studies on pathogenesis. It is useful to explore these referenced studies quantitatively and draw up a biological pathway. If appropriate a formal literature search may be undertaken to capture this literature comprehensively.
Experimental evidence	Does removing the presumed aetiological factor (eg, through treatment) eliminate the condition under investigation? This question can be addressed in randomised controlled trials. In the absence of such trials (whether or not randomised) uncontrolled comparisons may have to be included instead.

Detailed quality assessment is helpful in exploring heterogeneity and in generating inferences.[10] In this regard study design can be very informative. For example, when randomisation is not feasible, research may use cohort or case-control design. In studies on fetal exposure to potentially harmful maternal drugs,[15] these designs were used to evaluate causation. Dolovich et al[15] reported an increase in major fetal malformations after benzodiazepine exposure in utero in a meta-analysis of case controlled, but not of cohort studies (figure 1). The difference in results between cohort and case control is likely to be explained by recall bias or poor selection of control group, and weakens our inference about causation.[13 16]

Data synthesis

Formal examination of hypotheses concerning consistency, strength, temporality, dose-response relationship and the biological plausibility of suggested mechanisms facilitates the assessment of causality. Strength of association can be subdivided into magnitude (ie, how large was the effect) and precision (eg, p values and CI). For example, a meta-analysis of observational studies showed that bicycle helmets reduce the risk of head injuries in cyclists involved in a crash (OR 0.31, 95% CI 0.26 to 0.37).[17] The point estimate of OR and the upper limit of CI suggest a large, precise effect and, thus, a high degree of confidence in a causal association. In another example, the OR for exposure to benzodiazepines in pregnancy and association with major malformations in the newborn baby was 1.43 (0.89 to 2.31).[15] This result suggests a great deal of uncertainty in a causal association as the CI includes the possibility of no association at all. With regard to magnitude, ORs of above 2 and below 0.5 are often considered worthy of further causal exploration if derived from good-quality studies. In a randomised controlled trial (RCT) of folic acid supplementation starting before pregnancy, the association was large (RR 0.28, 95% CI 0.12 to 0.71).[18] When generated through randomised trials, such large effects are sufficiently compelling to lead to public policy changes. The precision achieved through meta-analysis does not by itself prove causation. One has to beware of the effect of bias and confounding. Even large effects from observational studies need careful consideration when inferring causation.[19]

Statistical analysis for heterogeneity and graphical representations may be used to explore consistency (figures 1 and 2). Statistical heterogeneity in the associations observed may arise when selected studies show qualitatively opposite results (eg, positive association as well as negative association seen through point estimates of ORs among cohort studies in figure 1). In this case the likelihood of there being a causal association will be low. Statistical heterogeneity in observed effects may also arise when selected studies show qualitatively same results (eg, positive association of different effect sizes). In this case, establishing a causal association may benefit from examination of a 'dose-response' relationship.

When feasible, combining study results (meta-analysis) may evaluate strength of association. Subgroup meta-analyses may be useful to explore the effect of study quality, temporality and dose-response relationships. In one example regarding the role of homocysteine in pre-eclampsia,[20] (figure 2) subgroup meta-analysis considering temporality of association reassured us, despite the strength of association among temporal studies being less than that overall, that the association met this causal criterion.

A biological gradient can be explored through meta-regression which investigates the effect of one or more study characteristics on the size of treatment effect, taking precision into account. Hooper et al[4] use meta-regression to show the association between modified dietary fat intake and cardiovascular disease. A genuine relation
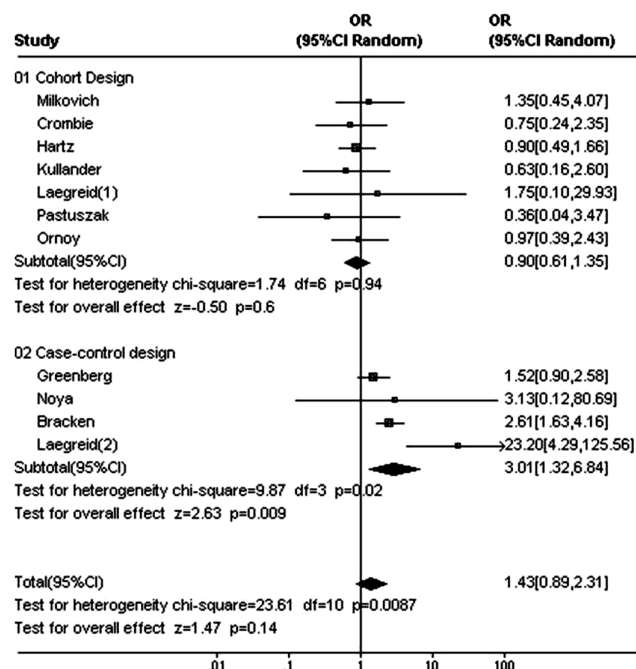
Figure 1 Quality assessment of studies: association of major malformations with prenatal exposure to benzodiazepines (adapted from[13 15]; values of point estimates of OR›1.0 indicate an association of malformations with exposure to benzodiazepines compared to no exposure).
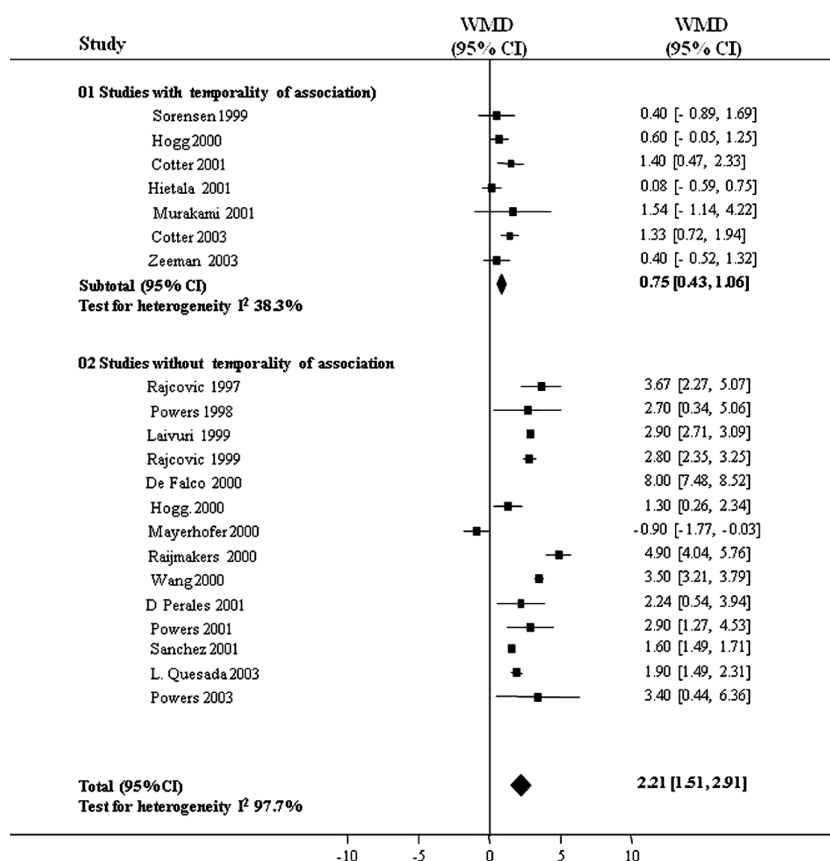
Figure 2 Temporality of association: subgroup meta-analysis on the role of homocysteine in preeclampsia (adapted from[20]; values of point estimates of weighted mean difference ›0 μmol/l indicate an association of hyper homocysteinaemia with preeclampsia).

Table 2 Dose-response relationship: pooled RRs (95% CI) for cardiovascular outcomes (number of pooled studies in parentheses after each effect estimate; heterogeneity statistic unavailable)[6]

Alcohol intake g/day versus none	Cardiovascular disease mortality (n=21 studies, n=1 184 956)
‹2.5	0.71 (0.57 to 0.89)[7]
2.5–14.9	0.77 (0.71 to 0.83)[15]
15–29.9	0.75 (0.70 to 0.80)[13]
30–60	0.85 (0.73 to 0.98)[10]
›60	0.99 (0.84 to 1.17)[6]

may be inferred when a slope is significantly different from zero. In another example for dose-response, (table 2) Ronksley et al[6] observed greater cardiac protection with increasing alcohol intake. In addition, when biological plausibility is demonstrated, for example, a favourable change in biomarkers of coronary heart disease (higher levels of high-density lipoprotein cholesterol and adiponectin and lower levels of fibrinogen) in those who drank moderate amounts of alcohol,[21] our confidence in a causal association is increased through existence of a disease mechanism.

Conclusion

The strength of the any causal inference that can be drawn from a review depends upon the rigour of the review methods employed, the validity of the review. This will be based on responses to several questions such as: was the initial search adequate; was the quality of the included studies adequate; were the findings both substantive and statistically significant; etc. Through systematic review both meta-analytic (quantitative) and criteria-based (qualitative) methods can be used together in making causal inferences.

When evaluating the effect of bias and confounding on causal criteria such as strength, consistency and dose-response of association, there is debate about how often there is correlation between results from RCTs and observational studies.[22 23] Conclusions stemming from the estimates of effect generated by the studies of different design being similar are valid only if the two groups of studies are similar in all respects other than the design itself. This is not often the case and medical advice based on observational studies has frequently been overturned when RCT evidence has emerged so that we now know we should rely on RCTs wherever feasible. For example, our belief about the protective effect of hormone replacement therapy on cardiovascular disease based on observational studies[24 25] has been overturned by RCT evidence.[26 27]

Traditionally, qualitative narrative review techniques are used to assess causal criteria. A systematic review, with judicious use of meta-analysis, provides an elegant solution to this difficult interpretative problem. When synthesising data to test hypotheses concerning causal criteria, there are no accepted scoring systems to decide whether there is sufficient evidence for causality or not. Ultimately, these assessments are judgments. But, a systematic approach to evidence synthesis and interpretation is likely to generate more transparent, robust inferences

than ad hoc considerations. When evidence is found to be lacking, reviews examining causation may be useful in identifying gaps in research.

Competing interests None.

References

1. Weed DL. On the use of causal criteria. *Int J Epidemiol* 1997;**26**:1137–41.
2. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**58**:295–300.
3. Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med* 1976;**49**:175–95.
4. Hooper L, Summerbell CD, Higgins JP, et al. Dietary fat intake and prevention of cardiovascular disease: systematic review. *BMJ* 2001;**322**:757–63.
5. Mignini LE, Villar J, Khan KS. Mapping the theories of preeclampsia: the need for systematic reviews of mechanisms of the disease. *Am J Obstet Gynecol* 2006;**194**:317–21.
6. Ronksley PE, Brien SE, Turner BJ, et al. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ* 2011;**342**:d671.
7. Wong SS, Wilczynski NL, Haynes RB, et al. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc* 2003:728–32.
8. Jacobson TZ, Duffy JM, Barlow D, et al. Laparoscopic surgery for pelvic pain associated with endometriosis. *Cochrane Database Syst Rev* 2009;**4**:CD001300.
9. Parascandola M, Weed DL. Causation in epidemiology. *J Epidemiol Community Health* 2001;**55**:905–12.
10. Weed DL. Meta-analysis and causal inference: a case study of benzene and non-Hodgkin lymphoma. *Ann Epidemiol* 2010;**20**:347–55.
11. Weed DL. Environmental epidemiology: basics and proof of cause-effect. *Toxicology* 2002;**181–182**:399–403.
12. Fox C, Mignini L, Khan K. Systematic reviews of research to assess causation: a guide to methods and application. *Eur Clinics Obstet Gynecol* 2006;**1**:251–6.
13. Khan K, Kunz R, Kleijnen J, et al. Systematic reviews to support evidence based medicine. Second edition. London: Hodder Arnold 2011.
14. Meads CA, Davenport CF. Quality assessment of diagnostic before-after studies: development of methodology in the context of a systematic review. *BMC Med Res Methodol* 2009;**9**:3.
15. Dolovich LR, Addis A, Vaillancourt JM, et al. Benzodiazepine use in pregnancy and major malformations or oral cleft: meta-analysis of cohort and case-control studies. *BMJ* 1998;**317**:839–43.
16. Khan KS, Wykes C, Gee H. Benzodiazepine use in pregnancy and major malformations or oral clefts. Quality of primary studies must influence inferences made from meta-analyses. *BMJ* 1999;**319**:919.
17. Thompson DC, Rivara FP, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database Syst Rev* 2000;**2**:CD001855.
18. Anonymous. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 1991;**338**(8760):131–7.
19. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. 1959. *Int J Epidemiol* 2009;**38**:1175–91.
20. Mignini LE, Latthe PM, Villar J, et al. Mapping the theories of preeclampsia: the role of homocysteine. *Obstet Gynecol* 2005;**105**:411–25.
21. Brien SE, Ronksley PE, Turner BJ, et al. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ* 2011;**342**:d636.

22. **Concato J**, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.

23. **Kunz R**, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–90.

24. **Stampfer MJ**, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;**20**:47–63.

25. **American College of Physicians**. Guidelines for counseling postmenopausal women about preventive hormone therapy. *Ann Intern Med* 1992;**117**:1038–41.

26. **Humphrey LL**, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002;**137**:273–84.

27. **Sare GM**, Gray LJ, Bath PM. Association between hormone replacement therapy and subsequent arterial and venous vascular events: a meta-analysis. *Eur Heart J* 2008;**29**:2031–41.