

Five steps to conducting a systematic review

Khalid S Khan, MB MSc, Regina Kunz, MD MSc,¹ Jos Kleijnen, MD PhD,² and Gerd Antes, PhD³

[Author information](#) ► [Copyright and License information](#) ►

This article has been cited by other articles in PMC.

Systematic reviews and meta-analyses are a key element of evidence-based healthcare, yet they remain in some ways mysterious. Why did the authors select certain studies and reject others? What did they do to pool results? How did a bunch of insignificant findings suddenly become significant? This paper, along with a book¹ that goes into more detail, demystifies these and other related intrigues.

A review earns the adjective systematic if it is based on a clearly formulated question, identifies relevant studies, appraises their quality and summarizes the evidence by use of explicit methodology. It is the explicit and systematic approach that distinguishes systematic reviews from traditional reviews and commentaries. Whenever we use the term *review* in this paper it will mean a *systematic review*. Reviews should never be done in any other way.

In this paper we provide a step-by-step explanation—there are just five steps—of the methods behind reviewing, and the quality elements inherent in each step (Box 1). For purposes of illustration we use a published review concerning the safety of public water fluoridation, but we must emphasize that our subject is review methodology, not fluoridation.

EXAMPLE: SAFETY OF PUBLIC WATER FLUORIDATION

[Go to:](#)

You are a public health professional in a locality that has public water fluoridation. For many years, your colleagues and you have believed that it improves dental health. Recently there has been pressure from various interest groups to consider the safety of this public health intervention because they fear that it is causing cancer. Public health decisions have been based on professional judgment and practical feasibility without explicit consideration of the scientific evidence. (This was yesterday; today the evidence is available in a York review^{2,3}, identifiable on MEDLINE through the freely accessible PubMed clinical queries interface [<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>], under 'systematic reviews'.)

STEP 1: FRAMING THE QUESTION

[Go to:](#)

The research question may initially be stated as a query in free form but reviewers prefer to pose it in a structured and explicit way. The relations between various components of the question and the structure of the research design are shown in [Figure 1](#). This paper focuses only on the question of safety related to the *outcomes* described below.



[Figure 1](#)

Structured questions for systematic reviews and relations between question components in a comparative study

Box 1 The steps in a systematic review

• Step 1: Framing questions for a review

The problems to be addressed by the review should be specified in the form of clear, unambiguous and structured questions before beginning the review work. Once the review questions have been set, modifications to the protocol should be allowed only if alternative ways of defining the populations, interventions, outcomes or study designs become apparent

• Step 2: Identifying relevant work

The search for studies should be extensive. Multiple resources (both computerized and printed) should be searched without language restrictions. The study selection criteria should flow directly from the review questions and be specified *a priori*. Reasons for inclusion and exclusion should be recorded

• Step 3: Assessing the quality of studies

Study quality assessment is relevant to every step of a review. Question formulation (Step 1) and study selection criteria (Step 2) should describe the minimum acceptable level of design. Selected studies should be subjected to a more refined quality assessment by use of general critical appraisal guides and design-based quality checklists (Step 3). These detailed quality assessments will be used for exploring heterogeneity and informing decisions regarding suitability of meta-analysis (Step 4). In addition they help in assessing the strength of inferences and making recommendations for future research (Step 5)

• Step 4: Summarizing the evidence

Data synthesis consists of tabulation of study characteristics, quality and effects as well as use of statistical methods for exploring differences between studies and combining their effects (meta-analysis). Exploration of heterogeneity and its sources should be planned in advance (Step 3). If an overall meta-analysis cannot be done, subgroup meta-analysis may be feasible

• Step 5: Interpreting the findings

The issues highlighted in each of the four steps above should be met. The risk of publication bias and related biases should be explored. Exploration for heterogeneity should help determine whether the overall summary can be trusted, and, if not, the effects observed in high-quality studies should be used for generating inferences. Any recommendations should be graded by reference to the strengths and weaknesses of the evidence

Free-form question

Is it safe to provide population-wide drinking water fluoridation to prevent caries?

Structured question

- *The populations*—Populations receiving drinking water sourced through a public water supply
- *The interventions or exposures*—Fluoridation of drinking water (natural or artificial) compared with non-fluoridated water
- *The outcomes*—Cancer is the main outcome of interest for the debate in your health authority
- *The study designs*—Comparative studies of any design examining the harmful outcomes in at least two population groups, one with fluoridated drinking water and the other without. Harmful outcomes can be rare and they may develop over a long time. There are considerable difficulties in designing and conducting safety studies to capture these outcomes, since a large number of people need to be observed over a long period. These circumstances demand observational, not randomized studies. With this background, systematic reviews on safety have to include evidence from studies with a range of designs.

STEP 2: IDENTIFYING RELEVANT PUBLICATIONS

[Go to:](#)

To capture as many relevant citations as possible, a wide range of medical, environmental and scientific databases were searched to identify primary studies of the effects of water fluoridation. The electronic searches were supplemented by hand searching of *Index Medicus* and *Excerpta Medica* back to 1945. Furthermore, various internet

engines were searched for web pages that might provide references. This effort resulted in 3246 citations from which relevant studies were selected for the review. Their potential relevance was examined, and 2511 citations were excluded as irrelevant. The full papers of the remaining 735 citations were assessed to select those primary studies in man that directly related to fluoride in drinking water supplies, comparing at least two groups. These criteria excluded 481 studies and left 254 in the review. They came from thirty countries, published in fourteen languages between 1939 and 2000. Of these studies 175 were relevant to the question of safety, of which 26 used cancer as an outcome.

STEP 3: ASSESSING STUDY QUALITY

[Go to:](#)

Design threshold for study selection

Adequate study design as a marker of quality, is listed as an inclusion criterion in Box 1. This approach is most applicable when the main source of evidence is randomized studies. However, randomized studies are almost impossible to conduct at community level for a public health intervention such as water fluoridation. Thus, systematic reviews assessing the safety of such interventions have to include evidence from a broader range of study designs. Consideration of the type and amount of research likely to be available led to inclusion of comparative studies of any design. In this way, selected studies provided information about the harmful effects of exposure to fluoridated water compared with non-exposure.

Quality assessment of safety studies

After studies of an acceptable design have been selected, their in-depth assessment for the risk of various biases allows us to gauge the quality of the evidence in a more refined way. Biases either exaggerate or underestimate the 'true' effect of an exposure. The objective of the included studies was to compare groups exposed to fluoridated drinking water and those without such exposure for rates of undesirable outcomes, without bias. Safety studies should ascertain exposures and outcomes in such a way that the risk of misclassification is minimized. The exposure is likely to be more accurately ascertained if the study was prospective rather than retrospective and if it was started soon after water fluoridation rather than later. The outcomes of those developing cancer (and remaining free of cancer) are likely to be more accurately ascertained if the follow-up was long and if the assessment was blind to exposure status.

When examining how the effect of exposure on outcome was established, reviewers assessed whether the comparison groups were similar in all respects other than their exposure to fluoridated water. This is because the other differences may be related to the outcomes of interest independent of the drinking-water fluoridation, and this would bias the comparison. For example, if the people exposed to fluoridated water had other risk factors that made them more prone to have cancer, the apparent association between exposure and outcome might be explained by the more frequent occurrence of these factors among the exposed group. The technical word for such defects is confounding. In a randomized study, confounding factors are expected to be roughly equally distributed between groups. In observational studies their distribution may be unequal. Primary researchers can statistically adjust for these differences, when estimating the effect of exposure on outcomes, by use of multivariable modelling.

Put simply, use of a prospective design, robust ascertainment of exposure and outcomes, and control for confounding are the generic issues one would look for in quality assessment of studies on safety. Consequently, studies may range from satisfactorily meeting quality criteria, to having some deficiencies, to not meeting the criteria at all, and they can be assigned to one of three prespecified quality categories as shown in [Table 1](#). A quality hierarchy can then be developed, based on the degree to which studies vary, but this was because randomized studies were non-existent and control for moderate quality and 18 of low quality.

Stumble



Study	Quality	Assessment
1	High	Randomized controlled trial
2	High	Randomized controlled trial
3	High	Randomized controlled trial
4	High	Randomized controlled trial
5	High	Randomized controlled trial
6	High	Randomized controlled trial
7	High	Randomized controlled trial
8	High	Randomized controlled trial
9	High	Randomized controlled trial
10	High	Randomized controlled trial
11	High	Randomized controlled trial
12	High	Randomized controlled trial
13	High	Randomized controlled trial
14	High	Randomized controlled trial
15	High	Randomized controlled trial
16	High	Randomized controlled trial
17	High	Randomized controlled trial
18	High	Randomized controlled trial
19	High	Randomized controlled trial
20	High	Randomized controlled trial
21	High	Randomized controlled trial
22	High	Randomized controlled trial
23	High	Randomized controlled trial
24	High	Randomized controlled trial
25	High	Randomized controlled trial
26	High	Randomized controlled trial
27	High	Randomized controlled trial
28	High	Randomized controlled trial
29	High	Randomized controlled trial
30	High	Randomized controlled trial
31	High	Randomized controlled trial
32	High	Randomized controlled trial
33	High	Randomized controlled trial
34	High	Randomized controlled trial
35	High	Randomized controlled trial
36	High	Randomized controlled trial
37	High	Randomized controlled trial
38	High	Randomized controlled trial
39	High	Randomized controlled trial
40	High	Randomized controlled trial
41	High	Randomized controlled trial
42	High	Randomized controlled trial
43	High	Randomized controlled trial
44	High	Randomized controlled trial
45	High	Randomized controlled trial
46	High	Randomized controlled trial
47	High	Randomized controlled trial
48	High	Randomized controlled trial
49	High	Randomized controlled trial
50	High	Randomized controlled trial
51	High	Randomized controlled trial
52	High	Randomized controlled trial
53	High	Randomized controlled trial
54	High	Randomized controlled trial
55	High	Randomized controlled trial
56	High	Randomized controlled trial
57	High	Randomized controlled trial
58	High	Randomized controlled trial
59	High	Randomized controlled trial
60	High	Randomized controlled trial
61	High	Randomized controlled trial
62	High	Randomized controlled trial
63	High	Randomized controlled trial
64	High	Randomized controlled trial
65	High	Randomized controlled trial
66	High	Randomized controlled trial
67	High	Randomized controlled trial
68	High	Randomized controlled trial
69	High	Randomized controlled trial
70	High	Randomized controlled trial
71	High	Randomized controlled trial
72	High	Randomized controlled trial
73	High	Randomized controlled trial
74	High	Randomized controlled trial
75	High	Randomized controlled trial
76	High	Randomized controlled trial
77	High	Randomized controlled trial
78	High	Randomized controlled trial
79	High	Randomized controlled trial
80	High	Randomized controlled trial
81	High	Randomized controlled trial
82	High	Randomized controlled trial
83	High	Randomized controlled trial
84	High	Randomized controlled trial
85	High	Randomized controlled trial
86	High	Randomized controlled trial
87	High	Randomized controlled trial
88	High	Randomized controlled trial
89	High	Randomized controlled trial
90	High	Randomized controlled trial
91	High	Randomized controlled trial
92	High	Randomized controlled trial
93	High	Randomized controlled trial
94	High	Randomized controlled trial
95	High	Randomized controlled trial
96	High	Randomized controlled trial
97	High	Randomized controlled trial
98	High	Randomized controlled trial
99	High	Randomized controlled trial
100	High	Randomized controlled trial

Table 1

Description of quality assessment of studies on safety of public water fluoridation

STEP 4: SUMMARIZING THE EVIDENCE

[Go to:](#)

To summarize the evidence from studies of variable design and quality is not easy. The original review² provides details of how the differences between study results were investigated and how they were summarized (with or without meta-analysis). This paper restricts itself to summarizing the findings narratively. The association between exposure to fluoridated water and cancer in general was examined in 26 studies. Of these, 10 examined all-cause cancer incidence or mortality, in 22 analyses. Of these, 11 analyses found a negative association (fewer cancers due to exposure), 9 found a positive one and 2 found no association. Only 2 studies reported statistically significant differences. Thus no clear association between water fluoridation and increased cancer incidence or mortality was apparent. Bone/joint and thyroid cancers were of particular concern because of fluoride uptake by these organs. Neither the 6 studies of osteosarcoma nor the 2 studies of thyroid cancer and water fluoridation revealed significant differences. Overall no association was detected between water fluoridation and mortality from any cancer. These findings were also borne out in the moderate-quality subgroup of studies.

STEP 5: INTERPRETING THE FINDINGS

[Go to:](#)

In the fluoridation example, the focus was on the safety of a community-based public health intervention. The generally low quality of available studies means that the results must be interpreted with caution. However, the elaborate efforts in searching an unusually large number of databases provide some safeguard against missing relevant studies. Thus the evidence summarized in this review is likely to be as good as it will get in the foreseeable future. Cancer was the harmful outcome of most interest in this instance. No association was found between exposure to fluoridated water and specific cancers or all cancers. The interpretation of the results may be generally limited because of the low quality of studies, but the findings for the cancer outcomes are supported by the moderate-quality studies.

RESOLUTION

[Go to:](#)

After having spent some time reading and understanding the review, you are impressed by the sheer amount of published work relevant to the question of safety. However, you are somewhat disappointed by the poor quality of the primary studies. Of course, examination of safety only makes sense in a context where the intervention has some beneficial effect. Benefit and harm have to be compared to provide the basis for decision making. On the issue of the beneficial effect of public water fluoridation, the review² reassures you that the health authority was correct in judging that fluoridation of drinking water prevents caries. From the review you also discovered that dental fluorosis (mottled teeth) was related to concentration of fluoride. When the interest groups raise the issue of safety again, you will be able to declare that there is no evidence to link cancer with drinking-water fluoridation; however, you will have to come clean about the risk of dental fluorosis, which appears to be dose dependent, and you may want to measure the fluoride concentration in the water supply and share this information with the interest groups.

The ability to quantify the safety concerns of your population through a review, albeit from studies of moderate to low quality, allows your health authority, the politicians and the public to consider the balance between beneficial and harmful effects of water fluoridation. Those who see the prevention of caries as of primary importance will favour fluoridation. Others, worried about the disfigurement of mottled teeth, may prefer other means of fluoride administration or even occasional treatment for dental caries. Whatever the opinions on this matter, you are able to reassure all parties that there is no evidence that fluoridation of drinking water increases the risk of cancer.

CONCLUSION

[Go to:](#)

With increasing focus on generating guidance and recommendations for practice through systematic reviews, healthcare professionals need to understand the principles of preparing such reviews. Here we have provided a brief step-by-step explanation of the principles. Our book¹ describes them in detail.

References

[Go to:](#)

1. Khan KS, Kunz R, Kleijnen J, Antes G. Systematic Reviews to Support Evidence-Based Medicine. How to Review and Apply findings of Health Care Research. London: RSM Press, 2003. [<http://www.rsmppress.co.uk/bkhan.htm>]
2. McDonagh M, Whiting P, Bradley M, *et al*. Systematic review of water fluoridation. *BMJ* 2000;321:855-9. [[PMC free article](#)] [[PubMed](#)]
3. NHS Centre for Reviews and Dissemination (CRD). A systematic review of water fluoridation. *CRD Report 18*. York: University of York, 2000 [<http://www.york.ac.uk/inst/crd/fluorid.htm>]