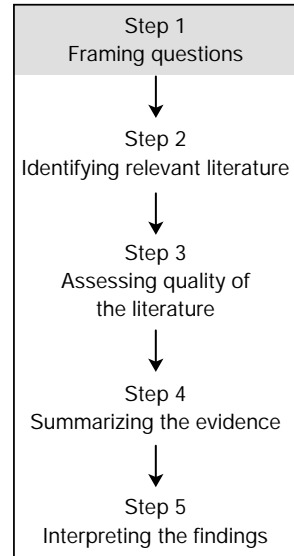


Step 1

Framing questions for a review

Systematic reviews are carried out to generate answers to focussed questions about health care and related issues. The key to a successful review project lies in the reviewer's ability to be precise and specific when stating the problems to be addressed in the review. This is a critical part of the review because, as will become apparent in subsequent steps, all other aspects of the review flow directly from the original questions. In this step, we will consider the question formulation process in detail and briefly look at the thinking required to examine the potential impact of variations in the different components of a review question.



1.1 An approach to formulating questions

Formulating questions is not as easy as it may sound. A structured approach to framing questions, which uses four components or facets, may be used. These components include the *populations*, *interventions* (or *exposures*), and *outcomes* related to the problem posed in the review, and the *designs* of studies that are suitable for addressing it. We can see the relationship between the various question components in a comparative study in Box 1.1.

After reading Box 1.1 the formulation of questions will probably seem like a daunting task to new reviewers. We may begin to have second thoughts, but we should not give up – help is at hand. This chapter of the book will take us through the question formulation process so that our review can have just the right start. It is well recognized that even quite experienced clinicians don't always find it easy to frame questions for evidence-based practice, so new reviewers can also expect to have a rough ride during the initial stages of their reviews. It will take some effort, but its value will be realized soon, as the rest of the review will flow directly and efficiently from the questions.

Most serious reviewers devote a substantial amount of time and effort in getting the questions right before

Question components

- The *populations*
- The *interventions*
- The *outcomes*
- The *study designs*

Free form question: It describes the query for which you seek an answer through a review in simple language (however vague).

Structured question: Reviewers convert free form questions into a clear and explicit format using a structured approach (see Box 1.2). This makes the query potentially answerable through existing relevant studies.

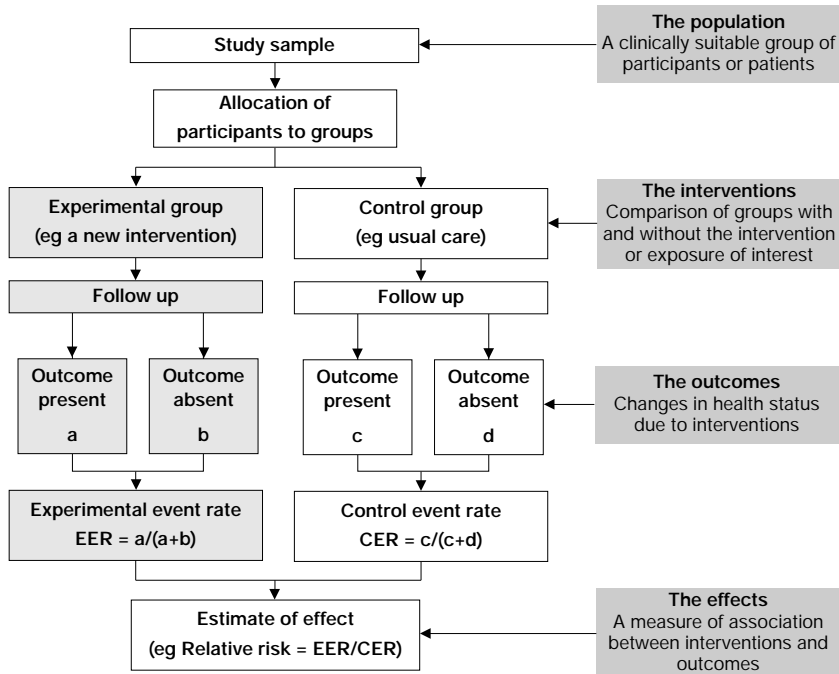
Box 1.1 Framing structured questions for systematic reviews

Question components

- The populations Succinct description of a group of participants or patients, their clinical problem and the healthcare setting.
- The interventions (or exposures) The main action(s) being considered, eg treatments, processes of care, social intervention, educational intervention, risk factors, tests, etc.
- The outcomes The clinical changes in health state (morbidity, mortality) and other related changes, eg health resource use.
- The study design The appropriate ways to recruit participants or patients in a research study, give them interventions and measure their outcomes.

Relationship between the question components in a comparative study

A comparative study assesses the effect of an *intervention* (or *exposure*) using comparison groups. For example, it may allocate participants or patients (with or without randomization) from a relevant *population* to alternative groups of *interventions* (or *exposures*) and follow them up to determine the effect of the *interventions* (or *exposures*) on *outcome*.



See related study designs in Box 1.4

embarking on a review. They do this because they want to avoid having to change questions later on during the review. We should make no exceptions. If there is any difficulty in figuring out the components of questions, we should first write them down in free form. We can then reconstruct the free form question into a structured format as exemplified in Box 1.2.

We should think of a *population* as a description of the group of participants or patients about whom evidence is being sought in the review. Imagine *interventions* as the actions or the alternatives being considered for the *population*. *Outcomes* are

Box 1.2 Some example questions

An example question about clinical effectiveness

Free form question: Which of the many available antimicrobial products improve healing in patients with chronic wounds?

Structured question

- The population In adults with various forms of chronic wounds in an ambulatory setting
- The interventions would systemic or topical antimicrobial preparations
- The outcomes improve wound healing?
- The study design A comparative study that allocates subjects with chronic wounds to alternative therapeutic interventions of interest and determines the effect of the interventions on wound healing (eg randomized controlled trial).

See case study 3 for a related review

An example question about aetiology

Free form question: Is exposure to benzodiazepines in pregnancy associated with malformations in the new-born baby?

Structured question

- The population In pregnant women. . . .
- The exposures does exposure to benzodiazepines during early pregnancy
- The outcomes cause malformations in the new-born baby?
- The study designs
 - A study that recruits women in early pregnancy, assesses their exposure to benzodiazepines, follows them up and examines their new-born babies to compare the rates of malformations among women with exposure and those without (cohort study).
 - A study that retrospectively compares exposure to benzodiazepines in early pregnancy among women who have given birth to a child with malformation with that among women who gave birth to a healthy child (case-control study).

See Box 5.2 for a related meta-analysis

continued

Box 1.2 Some example questions (continued)**An example question about test accuracy**

Free form question: Among postmenopausal women with abnormal vaginal bleeding, does pelvic ultrasound scan exclude uterine cancer accurately?

Structured question

- The population In postmenopausal women, within a community setting, with vaginal bleeding....
- The test does a uterine ultrasound scan test accurately predict
- The reference standard histological diagnosis of uterine cancer?
- The study design A study that recruits women from a relevant population, uses the test (scan) and a reference standard investigation to confirm or refute the presence of cancer (histology), and determines the accuracy with which the test identifies cancer (*see Box C4.3*).

See case study 4 for a related review

measures of what the *population* wants to achieve from the *interventions*, eg avoiding illness or death. Finally, we should think of how a study could be *designed* to examine the effect of our *interventions*. For example, by comparing *outcomes* between groups of a *population* with and without the *intervention*, the effect may be assessed in terms of illness avoided by use of an *intervention*.

The point about question formulation is that a structured approach should be used. The structure outlined in Box 1.1 should never become a ‘straight jacket’ and it may

This book will focus mainly on questions relating to **quantitative** effects of *interventions* (therapy, prevention, social care, etc.) or *exposures* (environmental agents, risk factors, etc.) in the context of **comparative study designs**.

be modified to meet the needs of our free form question depending on where our interest in health care lies. For example, in epidemiology the questions may be about aetiology. We can easily substitute the component *interventions* with *exposure* and frame the questions in terms of how *outcomes* might be different in *populations* exposed or not exposed to certain agents or risk factors (Box 1.2). For questions about the accuracy of screening or diagnostic tests we might substitute the component *intervention* with *test*, and *outcome* with *reference standard* against which the accuracy of the *test* will be measured (Box 1.2). In this way the proposed structure is versatile and adaptable for a wide range of question types.

1.2 Variations in *populations, interventions and outcomes*

Once the way in which questions are structured is understood (Box 1.1), we should be able to see that systematic reviews are analyses of existing studies within a given set

of *populations*, *interventions* and *outcomes*. We may have started with some scepticism about framing our question in this way; however, with the realization that different *populations*, *interventions* and *outcomes* exist within our free form question, we are likely to end up with many more than one question. If we have not, we should look hard to see if there is some variation within each one of our question components. This is critical – even in a straightforward question about antimicrobials for chronic wounds (Box 1.2), it should be clear that there are many types of chronic wounds (*populations*), antimicrobials (*interventions*) and ways of measuring wound healing (*outcomes*) (Box 1.3).

It is important to seriously consider how *populations*, *interventions* and *outcomes* might vary among existing studies. Such differences are important in defining study selection criteria (Step 2) and planning the tabulation of findings (Step 4). They are also relevant in understanding the reasons for variation in effects of *interventions* from study to study (Step 4) and in exploring the applicability of our findings (Step 5). Thus, conclusions of individual studies and reviews may vary depending on differences in the characteristics of their *populations*, the nature or delivery of their *interventions* and the types of *outcomes*. These issues are examined in detail later on in the book. Here we briefly examine their implications when framing questions.

Population characteristics may vary between studies with respect to patients' age and sex, severity of illness, presence of co-existing illnesses, etc. For instance, when the effect of home visits is studied among elderly people (Box 1.3), the *intervention* is effective among young-old rather than old-old people (Box 4.5). Similarly, the *intervention* features such as the care setting, compliance or intensity, additional routine care, etc. may also be associated with variable effects. For example, among elderly people, home visits are more effective if multidimensional assessments are used and follow-up is frequent (Box 4.5).

We need to identify all clinically relevant and important *outcomes*, which will help in examining the success or failure of our *interventions*. During our review it may become apparent that existing studies have not used *outcomes* we felt were relevant. Identification of these deficiencies in existing studies is important by itself, but sometimes when these data cannot be easily acquired, there may be a tendency to become interested in what are regarded as intermediate, surrogate or proxy *outcomes*. For example, when we are really interested in discovering the effect of fluoride therapy in preventing fractures, we might be tempted to investigate bone mineral content as a surrogate *outcome*, as it would be easier to obtain information about this. How misleading such an approach can be is demonstrated in a randomized controlled trial (*N Engl J Med* 1999; **322**: 802–9); bone density increased significantly (10–35% at different skeletal sites as compared to placebo) among the participants treated with fluorides, however, there was a nearly three-fold increase in non-vertebral fractures (control 24 vs fluorides 72, $p=0.01$), which was unexpected. This example makes it evident that conclusions from research based on surrogate *outcomes* are likely to be less valid for making decisions in practice.

When considering the *outcomes* for a review question, we should think about what we mean by health. Is it just the absence of illness or disease? This book mainly focuses on quantitative morbidity or mortality *outcomes*. It is becoming fashionable to

Box 1.3 Framing questions for reviews: Variations in *population, interventions, outcomes and study designs*

Two example questions about clinical effectiveness

Free form question: Which of the many available antimicrobial products improve healing in patients with chronic wounds?

Structured question (*expanded from Box 1.2*)

- | | | |
|---|--|---|
| <ul style="list-style-type: none"> • The population | <p>Adults with various forms of chronic wounds:</p> | <ul style="list-style-type: none"> • Diabetic ulcers • Venous ulcers • Pressure ulcers |
| <ul style="list-style-type: none"> • The interventions | <p>Antimicrobial preparations:
<i>versus</i>
Comparator:</p> | <ul style="list-style-type: none"> • Systemic preparations • Topical preparations • Other preparations |
| <ul style="list-style-type: none"> • The outcomes | <p>Clinical (various measures to quantify improvement in wound healing):</p> | <ul style="list-style-type: none"> • Complete healing • Wound area remaining • Healing scores |
| <ul style="list-style-type: none"> • The study design | <p>Experimental and observational studies:
(<i>see Box 1.4</i>)</p> | <ul style="list-style-type: none"> • Randomized controlled trials • Experimental studies without randomization • Cohort studies with concurrent controls |

See Case study 3 for a related review

Free form question: Do home visits improve the health of elderly people?

Structured question

- | | | |
|---|--|--|
| <ul style="list-style-type: none"> • The population | <p>Elderly people in various age groups:</p> | <ul style="list-style-type: none"> • Young-old • Middle age-old • Old-old |
| <ul style="list-style-type: none"> • The interventions | <p>Home visits:
<i>versus</i>
Comparator:</p> | <ul style="list-style-type: none"> • Intensive assessments • Frequent assessments • Usual care |
| <ul style="list-style-type: none"> • The outcomes | <p>Clinical (various measures to quantify health and health resource use):</p> | <ul style="list-style-type: none"> • Mortality • Functional status • Nursing home admissions |
| <ul style="list-style-type: none"> • The study design | <p>Experimental studies:
(<i>see Box 1.4</i>)</p> | <ul style="list-style-type: none"> • Randomized controlled trials • Experimental studies without randomization |

See Box 4.5 for a related meta-analysis

An example question about clinical and cost effectiveness

Free form question: To what extent is the risk of post-operative infection reduced by antimicrobial prophylaxis in patients undergoing hip replacement and is it worth the costs?

continued

Box 1.3 Framing questions for reviews: Variations in *population, interventions, outcomes* and *study designs* (continued)

Structured question

- | | | |
|---------------------|--|--|
| • The population | Patients undergoing hip replacement: | • Various types of procedures |
| • The interventions | Antimicrobial prophylaxis:
<i>versus</i>
Comparator: | • Various types of antibiotics
<i>versus</i>
• Placebo
• No antibiotics |
| • The outcomes | Clinical:
Economic: | • Post-operative infection
• Cost per infection prevented |
| • The study design | Clinical:

Economic: | • Experimental studies (<i>see Box 1.4</i>)
• Cost effectiveness analyses |

See Box 3.4 for related study quality assessment

consider the question of how to achieve optimal clinical *outcomes* with the smallest input of resources. This allows us to discover whether the investment in *interventions* is likely to be worthwhile. In this situation *outcomes* need to focus on the costs of providing health care in addition to clinical *outcomes* (Box 1.3). We will not cover these issues much beyond framing the questions.

1.3 Variations in *study designs*

Let us turn our attention to *study design*, the fourth component of a review question (Box 1.1). For a given set of *populations, interventions* and *outcomes*, reviews will provide summaries of existing studies that used different research *designs* (Box 1.2). Why is *design* so important? *Design* of a study determines the validity of the observed effects, ie our confidence that the results of a study are likely to approximate to the ‘truth’ for the participants or patients studied depends on the soundness of its *design*. In this way *design* serves as a marker of study quality. Its importance cannot be emphasized enough. Ultimately the strength of a review’s inferences depends on the integrity of *designs* of the available studies.

Some reviewers consider certain *study designs* to be superior because they feel that the *design* has an inherent value in itself. For example, they may focus exclusively on randomized studies when conducting reviews. Such a view ignores the fact that addressing different types of questions may require the use of different *study designs*. As an example, a question about accuracy of a *test* would require a *study design* that prospectively (without randomization)

Valid results are said to be unbiased. **Bias** either exaggerates or underestimates the ‘true’ effect of an *intervention* or *exposure*.

The **quality** of a study depends on the degree to which its design, conduct and analysis minimizes **biases**.

recruits all eligible patients, employs the *test* and the *reference standard* investigation to confirm or refute the presence of disease, and determines the accuracy with which the test correctly identifies disease (case study 4). Assessment of long-term or rare *outcomes*, particularly when examining the safety of *interventions* (as in case study 2), would be more suited for an observational *design*, not an experimental study. For example, cohort and case-control studies, not randomized trials, would evaluate the effect of *exposure* to benzodiazepines in pregnancy on rare malformations in the newborn baby (Box 5.2).

Effectiveness is the extent to which an *intervention* (therapy, prevention, diagnosis, screening, education, social care, etc.) produces beneficial *outcomes* under ordinary day-to-day circumstances.

Even for questions concerning effectiveness of *interventions*, where randomized trials are generally preferred, it might be difficult to justify a restriction to using randomized studies only. This may be particularly true when such studies are unethical. Sometimes there is just a dearth of randomized studies. For example, in the review on antimicrobials for chronic wounds (case study 3), despite a comprehensive search, only four clearly randomized studies could be found, so other *designs* had to be included. On the other hand, in case study 2, where the review considered safety of water fluoridation, no randomized studies had been published, so it became necessary to consider various other *designs*. Sometimes a review may consider a number of separate but related questions. For example, if a review is to include an assessment of efficiency in addition to effectiveness, then *study designs* for economic evaluation will also be required (Box 1.3). Thus, it might be necessary to consider different *designs* simultaneously in some review questions. This multiplicity of *designs* has implications for study quality assessment (Step 3) and synthesis (Step 4).

Efficiency (cost effectiveness) is the extent to which the balance between input (costs) and output (*outcomes*) of *interventions* represents value for money.

Insistence on randomized studies, ignoring other types of evidence, might paralyse reviewers as such reviews might never find any studies. When faced with having to make decisions for practice, using the best available evidence is likely to be better than not using any evidence at all. We will need to explore the nature of our questions (effectiveness, aetiology, efficiency, accuracy, etc) and the different ways of addressing the specific issues before us, ie *populations*, *interventions* and *outcomes*. Then we should select the *study designs* that are likely to provide the most valid answers and

develop a hierarchy of *study designs* suitable for our review. This approach will help us define inclusion and exclusion criteria for selecting studies of a minimum acceptable quality (Step 2).

Each question type has a *design* hierarchy of its own. In this book we focus mainly on questions relating to health effects of *interventions* and *exposures*. These questions usually focus on how one *intervention* or *exposure* compares with another. A hierarchy of *designs* for studies addressing such issues is shown in Box 1.4. The most sound

Box 1.4 A hierarchy of *study designs* for questions about effectiveness of healthcare interventions

Description of the <i>design</i>	Levels assigned to evidence based on soundness of <i>design</i> [†]
<p>Experimental study A comparative study* in which the use of different <i>interventions</i> among participants is allocated by the researcher.</p> <ul style="list-style-type: none"> • Randomized controlled trial (with concealed allocation) Random allocation of participants to an <i>intervention</i> and a control (eg placebo or usual care) group, with follow-up to examine differences in <i>outcomes</i> between the two groups. Randomization (with concealment of allocation sequence from caregivers) avoids bias because both known and unknown determinants of outcome, apart from the intervention, are usually equally distributed between the two groups of participants. 	I
<ul style="list-style-type: none"> • Experimental study without randomization (sometimes erroneously called quasi-experimental or quasi-randomized or pseudorandomized studies) A study in which the allocation of participants to different <i>interventions</i> is managed by the researcher but the method of allocation falls short of genuine randomization, eg alternate or even-odd allocation. Such methods fail to conceal the allocation sequence from caregivers. 	
<p>Observational study with control group A comparative study* in which the use of different <i>interventions</i> among participants is not allocated by the researcher (it is merely observed).</p> <ul style="list-style-type: none"> • Cohort study Follow-up of participants who receive an <i>intervention</i> (that is not allocated by the researcher) to examine the difference in <i>outcomes</i> compared to a control group, eg participants receiving no care. • Case-control studies Comparison of <i>intervention</i> rates between participants with the outcome (cases) and those without the <i>outcome</i> (controls). 	II [‡]
<p>Observational study without control groups</p> <ul style="list-style-type: none"> • Cross-sectional study Examination of the relationship between <i>outcomes</i> and other variables of interest (including interventions) as they exist in a relevant <i>population</i> at one particular time. • Before-and-after study Comparison of <i>outcomes</i> in study participants before and after an <i>intervention</i>. • Case series Description of a number of cases of an <i>intervention</i> and their <i>outcomes</i>. 	III
<p>Case reports Pathophysiological studies or bench research Expert opinion or consensus</p>	IV

* A comparative study assesses the effect of an intervention using comparison groups. See Box 1.1 for an example flow chart of such a study

[†] See Box 5.4 for use of levels of evidence in grading recommendations for practice based on reviews of effectiveness

[‡] In Level II evidence, experimental studies without randomization (and allocation concealment) are considered better than cohort studies, which in turn are considered better than case-control studies

study design in this context (often called Level I evidence) is one that randomly allocates (concealing the assignment code) participants from a relevant *population* to the alternative *interventions* of interest. This *design* serves to remove selection bias and when conducted well such studies rank at the top of the *study design* hierarchy for effectiveness evidence. Next in the hierarchy are Level II studies where there are three sub-levels. Experimental studies where the allocation of participants or patients is controlled by the researcher, but falls short of genuine randomization and allocation concealment, are considered better than cohort studies, which in turn are considered better than case-control studies. As indicated above, for many reviews experimental studies will not exist (case study 2) or they might be scarce (case study 3). Hence, reviews may have to be conducted using studies of an inferior *design* or using studies with a mixture of *designs*. If our review has several *study designs*, it would be prudent to carefully plan quality assessments (Step 3), stratify study synthesis by *design* and quality (Step 4) and interpret findings cautiously, relying on methodologically sound studies (Step 5).

1.4 Modification of questions during a review

It is important that review questions are formulated *a priori*, that is before the review work is actually commenced. Otherwise the review process may be unduly driven by presuming particular findings. In order to get the questions correct at the beginning, it may be worth involving experienced reviewers and practitioners in the process. This is just one of several reasons why it is considered unwise to prepare a review alone.

Questions will initially be developed without detailed knowledge of much of the relevant literature. Therefore, we should not be surprised if it becomes evident during the review that some questions need to be modified in light of the accumulated research. The commandment ‘thou shall pose questions for a review *a priori*’ should not be applied too rigidly. We should allow exploration of unexpected issues in the review process; as a greater understanding of the problem is developed during the course of the review, it would be foolish not to do so. If the ongoing work identifies a need for answering questions that had not been foreseen, it would be quite reasonable to raise new questions or to modify existing questions. Such modifications are justifiable if they are based on the realization of alternative ways of defining the *populations*, *interventions*, *outcomes* or *study designs*, which were not considered earlier.

Revision of questions will inevitably have some implications for the review work. The protocol would have to be revised. Literature searches (Step 2), which are usually conducted before questions are refined, may also need refinement and they might have to be run again in the light of the changes to the questions. Study selection criteria will have to be altered. For example, in the review of safety of water fluoridation in case study 2, the original questions were modified in the light of information gathered about the extent and range of quality of available evidence during the initial part of the review. This led to changes in study selection criteria, which are provided in detail in the published report of the review (www.york.ac.uk/inst/crd/fluorid.pdf). Reviewers

should not be economical with the truth about question formulation and refinement. It is helpful to be explicit about the modifications and indicate which questions were posed *a priori* and which were generated during the review work.

Summary of Step 1: Framing questions for a review

Key points about appraising review articles

- Examine the abstract and the introduction to see if the review is based on predefined questions.
- Examine the methods and other sections to check if questions were modified during the review process.
- Can we be sure that the questions have not been unduly influenced by the knowledge of results of the studies?

Key points about conducting reviews

- The problems to be addressed by the review should be specified in the form of clear, unambiguous questions before beginning the review work.
- Questions should be structured, eg in terms of *population*, *interventions*, *outcomes* and *study designs* relevant to the healthcare issues being addressed in the review.
- Characteristics of the *populations*, differences in *interventions*, variation in *outcomes* and variety in *study designs* may influence the results of a review. The impact of these factors should be carefully considered at this stage.
- Once the review questions have been set, modifications to the protocol should only be allowed after careful consideration. Sometimes, alternative ways of defining the *populations*, *interventions*, *outcomes* or *study designs* become apparent after commencing the review. In this situation it would be reasonable to alter the original questions, but these modifications should not be driven by the knowledge of results of the studies.

