

WHO Handbook for Guideline Development

March 2010

The handbook is frequently updated.. Your feedback is appreciated. Please send suggestions to grcinfo@who.int.

Contents

Introduction	4
Definitions.....	4
The WHO guideline approval process.....	8
Approval for development.....	8
Final approval.....	9
Guideline development process.....	11
Planning and defining the scope of a guideline.....	12
Before you start	12
Practical planning	13
Defining the scope of the guideline.....	14
Guideline development group - function and composition	17
Purpose of the group	17
Composition of the group.....	17
Declaration and management of conflicts of interests (COI)	20
1. Assessment prior to final selection of group members.	20
2. Declaration.....	22
3. Reporting.....	23
What to do when you have too many conflicts? Can participants with conflicts attend as 'observers'?	23
Formulating questions and choosing outcomes	25
Choosing and rating outcomes	28
Evidence retrieval, assessment and synthesis.....	30
Prioritizing evidence retrieval.....	31
Evidence retrieval	32
Evidence assessment	34
Grading the quality of evidence and the strength of recommendations.....	37
Evaluating the quality of evidence	38
Summarizing findings.....	39
Using GRADE tables	40
Assessing evidence for other types of questions.....	45
Assessing cost and resource implications.....	47
What is it about?.....	47
What do I need to do?	48
What can be done?	49
Incorporating values and preferences	51
Formulating recommendations.....	52
How a guideline development group decides on recommendations	52
Grading recommendations.....	53
Research needs and priorities.....	55
Peer review and plans for updating	56
The peer review process.....	56
Length of validity of guidelines.....	56
Producing and publishing your guideline	58
Guideline format.....	58

Implementation and evaluation of impact	60
Implementation	60
Evaluation and monitoring	60
APPENDICES	62
Appendix I. Effective guideline development group meetings	63
Appendix II. Critical appraisal of systematic review sheet	63
Appendix III. The PRISMA criteria for reporting of systematic reviews.....	66

Introduction

All publications containing WHO recommendations must be approved by WHO's Guidelines Review Committee (GRC). Such publications are required to meet an unmet need, to be up-to-date, and to be developed using internationally accepted best practices, including the appropriate use of evidence. This handbook provides guidance on the development of documents or publications containing WHO recommendations, and sets out the procedures to follow when such a document is submitted to the GRC for approval. To facilitate ease of reading, the term "guideline" is used to refer to any document containing WHO recommendations.

The GRC reviews every WHO guideline twice during its development – once after the scope of the guideline has been defined at the initial planning stage (initial approval for development), and again after the recommendations have been developed and before the document is edited and published (final approval).

The GRC meets monthly to review submissions. To allow adequate time for review, all relevant documents must be submitted to the GRC no later than two weeks before the date of the next meeting.

The GRC can be contacted at grcinfo@who.int. To download the relevant documents needed for GRC clearance, please go to <http://intranet.who.int/homes/rpc/grc/>. The forms are :

[planning clearance form](#)

[planning clearance checklist](#)

[executive clearance form](#)

[executive clearance checklist](#)

Definitions

A WHO guideline is any document, whatever its title, that contains WHO recommendations about health interventions, whether they be clinical, public health or policy interventions. A recommendation provides information about what policy-makers, health-care providers or patients should do. It implies a choice between different interventions that have an impact on health and that have ramifications for the use of resources.

Information products that are NOT considered guidelines include:

- documents containing standards for manufacturing health technologies, such as pharmaceuticals and vaccines;
- “how to” documents, or operational manuals (e.g. how to set up a research project or how to implement a service);

- documents that describe standard operating procedures for organizations or systems;
- documents that state established principles (e.g. ethics, human rights, WHO constitutional issues);
- documents that provide information on different options for interventions without recommending any particular intervention.

If you are not sure whether your proposed document is a guideline, please submit it to the GRC for review.

If you are planning to produce a guideline, consider which of the following types of product best fits your purpose. The type of product will determine the methods and timeframe for development.

Rapid advice guidelines

A “rapid advice” guideline is produced in response to a public health emergency (e.g. the emergence of a new SARS-type outbreak) in which WHO is required to provide rapid global leadership and guidance. This type of document needs to be produced quickly (in about 1–3 months) and will be evidence-informed, but it may not be supported by full reviews of the evidence. It will be prepared mainly by the responsible WHO staff members with external consultation and peer review. It should be published with a “review-by” date that indicates when the guidance will become invalid, or when it will be updated or converted to a standard guideline.

Standard guidelines

A “standard” guideline is produced in response to a request for guidance in relation to a change in practice or controversy in a single clinical or policy area – such as treatment of postpartum haemorrhage or avian influenza, or minimum requirements for safe delivery of HIV care. A standard guideline is not expected to cover the full scope of the condition or public health problem. This guideline will usually take 9–12 months to complete and should be prepared after consultation on the scope of the guideline and the issue that it covers. It should be supported by systematic evidence reviews (that could be commissioned externally) and one or two meetings of the guideline development group for consultation. A standard guideline may have a specified “use-by” date depending on the expected rate of change of evidence in the topic area. Most WHO guidelines fall into this category.

Full guidelines

A “full” or “management” guideline is one that provides complete coverage of a health topic or disease, such as dengue fever. It would be expected to include recommendations in relation to all aspects of the topic (e.g. surveillance, diagnosis, public health and clinical interventions) and to be fully based on evidence reviews. It is likely to take 2–3 years to complete, will require several meetings of a guideline development group, and should therefore be prepared only when WHO is the most appropriate agency to undertake the task or when there is likely to be no other group producing the guideline.

Compilation of guidelines

A compilation of guidelines contains current recommendations from WHO and other sources, but does not include any new recommendations. Compilations of guidelines are subject to GRC review. All recommendations included must be current and should be referenced thoroughly and accurately. Producing a compilation of guidelines can be complex and updating may be difficult since individual recommendations may go out of date at different times.

In principle, all recommendations used in a compilation should be updated by WHO. However, recognizing that WHO resources are limited, this may not be realistic. Members of the guideline development group should discuss and agree on an acceptable level of quality and document their decisions carefully. It is strongly recommended to use the [AGREE instrument](#) to do this.

It is also important that recommendations used in a compilation are of adequate quality. WHO recommendations are considered of adequate quality for use in a compilation if they were cleared by the GRC from 2009 onwards. If compiled recommendations have not been cleared by the GRC, an explicit and systematic process must be in place to ensure the quality of the compiled guidance. Production times for compilations of guidelines vary widely.

Some guideline compilations do not require GRC review. These are:

- documents in which all the recommendations have previously been cleared by the GRC under its full (not transitional) requirements;
- documents that are clearly limited to operational guides for such guidelines.

Guideline compilations that do require GRC review are documents in which all the recommendations were initially published without GRC review, regardless of whether they were published by WHO or another organization. Submission to the GRC is required for assessment of the quality of the recommendations and the use of evidence.

Adaptation of guidelines

Recommendations from various sources used as the basis for developing guidance in a different setting (e.g. maternal health recommendations in emergency settings) are considered adaptations of existing guidelines. Adaptations of guidelines must follow standard GRC procedures.

Guidelines prepared in collaboration with other organizations

Guidelines for the management of clinical conditions are produced by many organizations, including national agencies and specialist medical societies. From time to time, it may be appropriate for WHO to collaborate with these groups to produce a joint guideline. However, national agency guidelines usually have a much narrower focus than those from WHO, and international society guidelines may have inherent problems owing to conflicts of interest in the funding of their development. The GRC will make case-by-case assessments of these types of proposal. However, joint

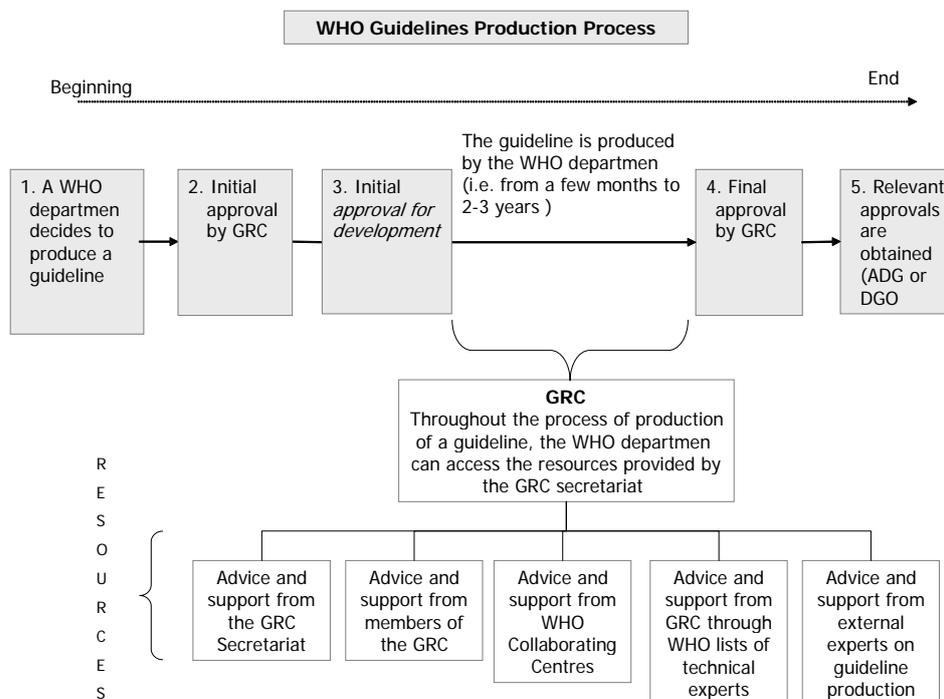
guidelines must follow current WHO guideline development standards as outlined in this handbook. In addition to being aware of potential problems with regard to copyright and ownership, it is important to note that:

- adaptation or endorsement of another organization's guideline should be initiated by the WHO department concerned and not by the external group;
- adaptation or endorsement of another organization's guideline can be considered when no WHO guideline exists or an existing WHO guideline is outdated;
- minimum standards for WHO guidelines should be met (no funding from commercial sources, conflicts of interest declared and reported)
- the approach to reviewing and summarizing evidence should be consistent with that recommended for WHO guidelines;
- WHO should ensure global representation of experts in the development of the recommendations;
- the recommendations should be appropriate for a global audience.

The WHO guideline approval process

The overall WHO guideline approval process is shown in Figure 1.

Fig. 1. Approval process for WHO guidelines



Approval for development

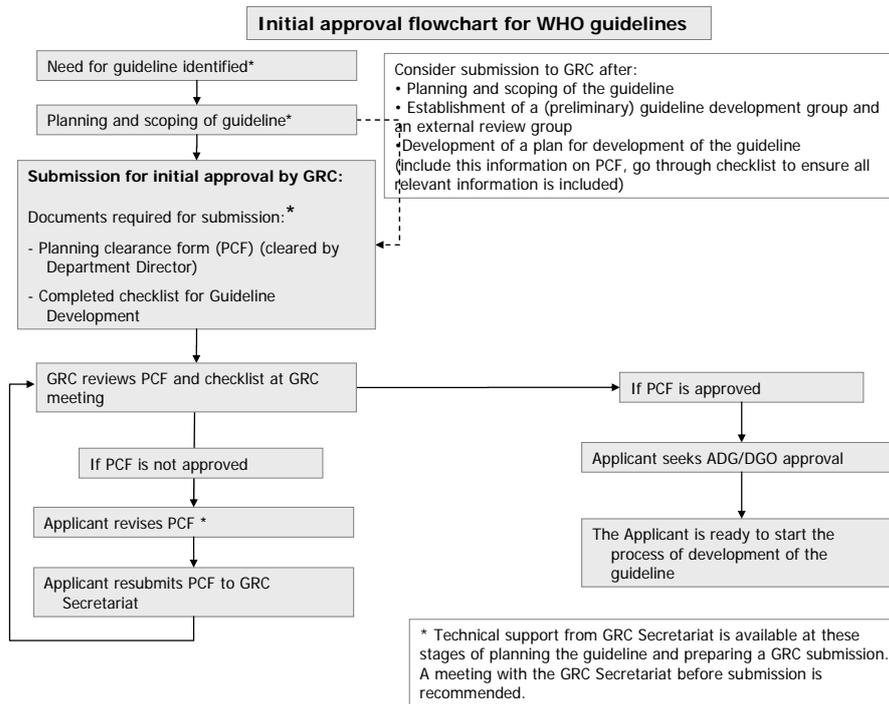
Initial approval must be obtained from the GRC before development of the guideline begins.

The plan for guideline development is usually submitted to the GRC after the scope of the guideline has been defined. Before submission, you should have:

- identified any existing related guidance;
- defined the scope of the guideline, including objectives, target audience and draft research questions;
- established the WHO guideline steering group;
- established a (preliminary) guideline development group and an external review group;
- outlined a plan and timeline for development of the guideline.

This information should be included in the planning clearance form (and in annexes, if necessary). Once planning clearance is provided, the guideline should be developed according to the procedures outlined here. Figure 2 shows the initial guideline approval process.

Fig. 2. Flowchart showing the initial process for guideline approval

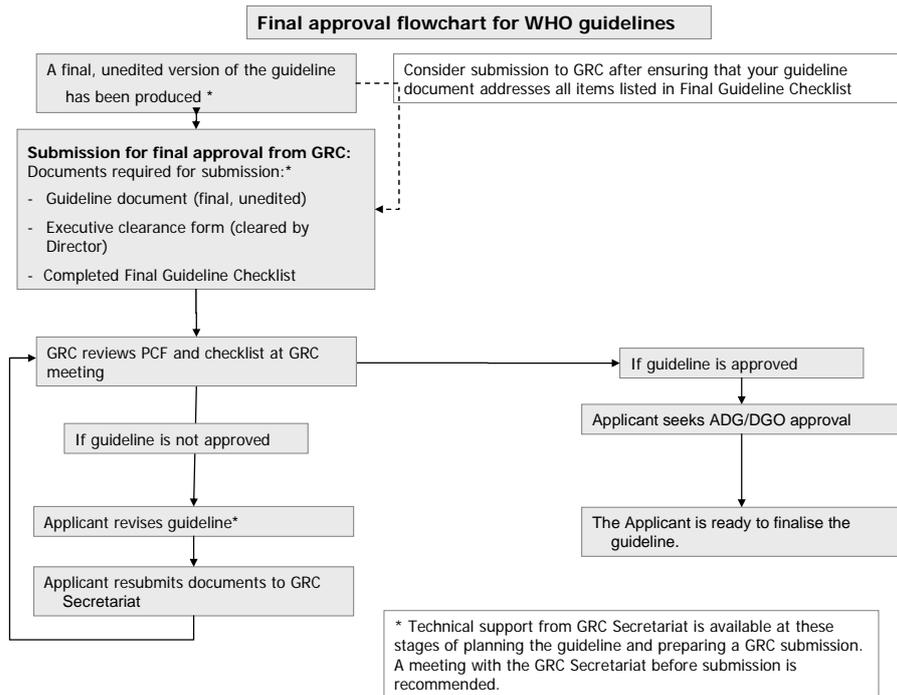


Final approval

After the guideline has been developed according to the process outlined here, approval of the guideline must be obtained from the GRC before the document is finalized, edited and published.

The final document should address all the items in the final guideline checklist (see <http://intranet.who.int/homes/rpc/grc/guidelines/>). Figure 3 shows the final approval process.

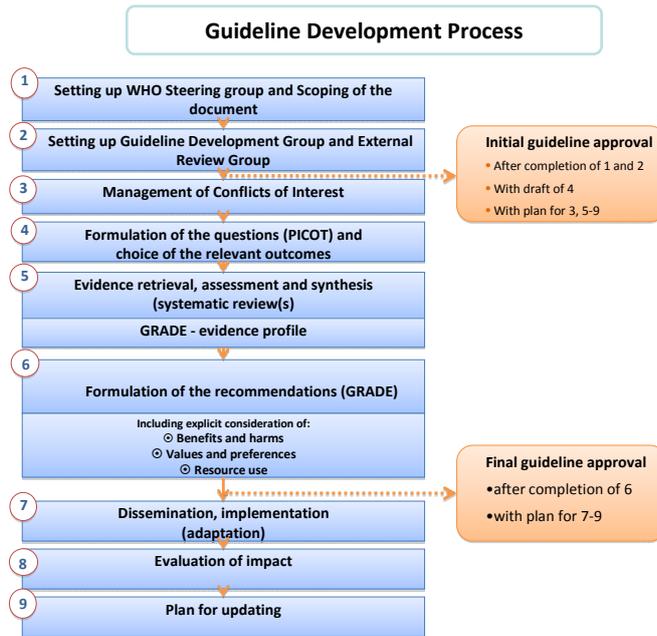
Fig. 3. Flowchart showing the final guideline approval process



Guideline development process

The process for WHO guideline development is shown here in Figure 4.

Fig. 4. The WHO guideline development process



Planning and defining the scope of a guideline

Before you start

Before starting to develop a guideline, it is important to consider a number of questions which are dealt with below.

Consider why the guideline is necessary

Who is asking for this guideline? Is it a request from one or more WHO member states? WHO guidelines generally should meet a global need, have a public health perspective and not duplicate existing resources. If an existing guideline meets the need, a new one is not required.

Why does this guideline need to be developed by WHO now? Is it required by the organization's governing bodies? Are there already guidelines on the same topic from other organizations or other WHO departments? Is the best advice on this topic available only from WHO?¹

Is the development of the guideline part of a departmental programme of work? Implementation of a guideline by WHO headquarters or by countries will be much easier if it fits with a programme or project. If no programme or project exists, is it really necessary to prepare the guideline? Who is likely to implement it? If you cannot identify a process for implementation, then you should not start.

Is the guideline intended to respond to poor practice or to try to change clinical practice or health policy? This should be the focus of most guidelines, and it is what differentiates guidelines from textbooks or reference works. However, development of the guideline is only one step in the process of implementing change in practice.

Is the guideline a response to a situation where need for advice is urgent? If so, see the description of rapid advice guidelines above. These guidelines usually need to be produced and published as quickly as possible, ideally in 1–3 months, and therefore the requirements and processes are different from those of other guidelines.

Do you have agreement from your director and Assistant Director-General? You will need to have formal approval from both, so agreement at the outset is important.

Are there other departments that should be involved, or that might be producing similar products? The answer to this is nearly always yes. Avoid duplication of existing guideline development efforts. Early in the process, consult other relevant WHO departments, the GRC secretariat or the WHO library to help you find published works that are relevant to your planned guideline. Decide which department has primary responsibility for the guideline and who will be involved in developing it.

If you cannot answer all these questions, it is probably best not to start.

¹ Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development. 2. Priority setting. *Health Research Policy and Systems* 2006;4:14 (available at <http://www.health-policy-systems.com/content/4/1/14>).

Practical planning

Further questions that will need to be answered are summarized here.

Who is your target audience?

Most WHO guidelines need to address multiple audiences, which makes them challenging to produce. If you can identify the key target audience, your task will be easier and may meet their needs. Writing documents to meet the needs of policy-makers, health-care managers and clinicians simultaneously is not straightforward and should be avoided wherever possible.

When does the guideline need to be completed?

Realistically, producing a good quality guideline will take at least 9–12 months if all the evidence has already been synthesized and you have someone to write it. If the guideline is going to cover a large number of questions, it may take up to 2–3 years to produce. Again, consider whether you really need to do this.

How much money is available?

For a standard WHO guideline – assuming that you will need to contract an individual or group to prepare evidence summaries (based on existing systematic reviews), hold a single consultation meeting, pay for writing and editing and a small print run of the final document – a reasonable cost estimate is US\$ 100 000. If you do not have this amount of funding, or are not sure it will be made available, do not begin. Note that WHO may not accept money for guideline development from commercial bodies or from professional organizations sponsored by commercial bodies. Sources of funding for guidelines may need to be approved by the WHO legal office.

Are there existing guideline documents that cover the same issue?

If so, what is the added value and justification for the proposed document? If the existing guidelines are from high-quality national authorities (e.g. the UK's National Institute for Health and Clinical Excellence) and a WHO version will truly provide added value, the existing guidelines can be used as a starting point. However, specialist society guidelines need to be treated with much more caution as they are often funded largely by commercial entities, without consideration of conflicts of interest. Consider updating existing WHO recommendations if they are out of date or of low quality.

If you wish to start with existing guidelines, assess their quality first. One option is to use the [AGREE instrument](#).

What existing scientific evidence can guide the recommendations?

Do you know of existing systematic reviews? If not, it is worth doing a preliminary literature search at this stage to get a sense of what information is available. For standard and full WHO guidelines, a systematic search for evidence should be

completed before developing the recommendations. If there is no evidence, what will be the basis of your guideline?

Who should be consulted and involved?

It is worth spending some time at the beginning of the process to draw up a list of key external organizations, experts and stakeholders who will need to be consulted or involved in the process. First, identify your WHO guideline steering group (WHO staff members responsible for guideline development). Second, identify members of your guideline development group who will be actively involved in the development of the guidelines (ideally 15–20 persons). Third, there should usually be an additional group of experts or organizations serving on an external review group that you may wish to consult on the scope of the document, the questions it covers and the choice of important outcomes for decision-making, as well as to review the completed draft guideline. The external review group may include groups likely to oppose or criticize the output on the basis of scientific or philosophical differences. While it may not be possible to reach agreement with them, it is important to consider their input. In addition, many of these groups and experts will play a key role in implementation of the recommendations in the guideline; they are more likely to help implement the recommendations if they are involved from the beginning.

In what format(s) will your guideline be produced?

Printed documents may not be the most useful format. Electronic versions may be more practical and cheaper, perhaps accompanied by short paper publications, wall charts, pamphlets, etc.

Are you planning translations?

If so, in which languages? Consider the implications of translations for the budget and time frame, since translations can be expensive and time-consuming.

Defining the scope of the guideline

Defining the scope of the guideline refers to the process of defining the content, clinical questions and likely recommendations. If you get this stage right, the later stages will be much easier.

The process is as follows:

1. Convene a small group of WHO staff to define the scope of the guideline, including representatives of all relevant departments (the WHO steering group).
2. List the priority topics for the guideline. If you are working on a clinical area (e.g. dengue fever) what must be included? Resist the temptation to write a textbook. Concentrate on the interventions or policies where change in practice is desired, and areas where there is controversy. Also consider the feasibility of implementing potential recommendations. Although some background information may be useful, try to avoid repeating standard information (e.g.

epidemiology, pathology, pharmacology) on the topic unless this is the area of controversy you wish to resolve in the guideline.

3. If you know what recommendations need to be formulated, list these as well.
4. Do you need to include all of these topics, questions and recommendations? The group should try to restrict the final list to the minimum at this stage, as it will inevitably expand during the development of the guideline.
5. On the basis of the topic list and possible recommendations, formulate the key questions to be answered in the guideline. These questions will guide the evidence synthesis. It is useful to formulate the questions using the PICOT format (see section on [formulating questions](#) below).
6. Once you have a complete scope that is agreed by the group, it should be circulated to the external review group for comments. (They should be reminded that WHO is producing a guideline, not a textbook, as the responses will almost always tend to expand the planned scope.)
7. Once you have the feedback, do a reality check. Is what you are trying to do feasible? Is your time frame reasonable? Do you have the money to do it?

Defining the scope of an update of a guideline

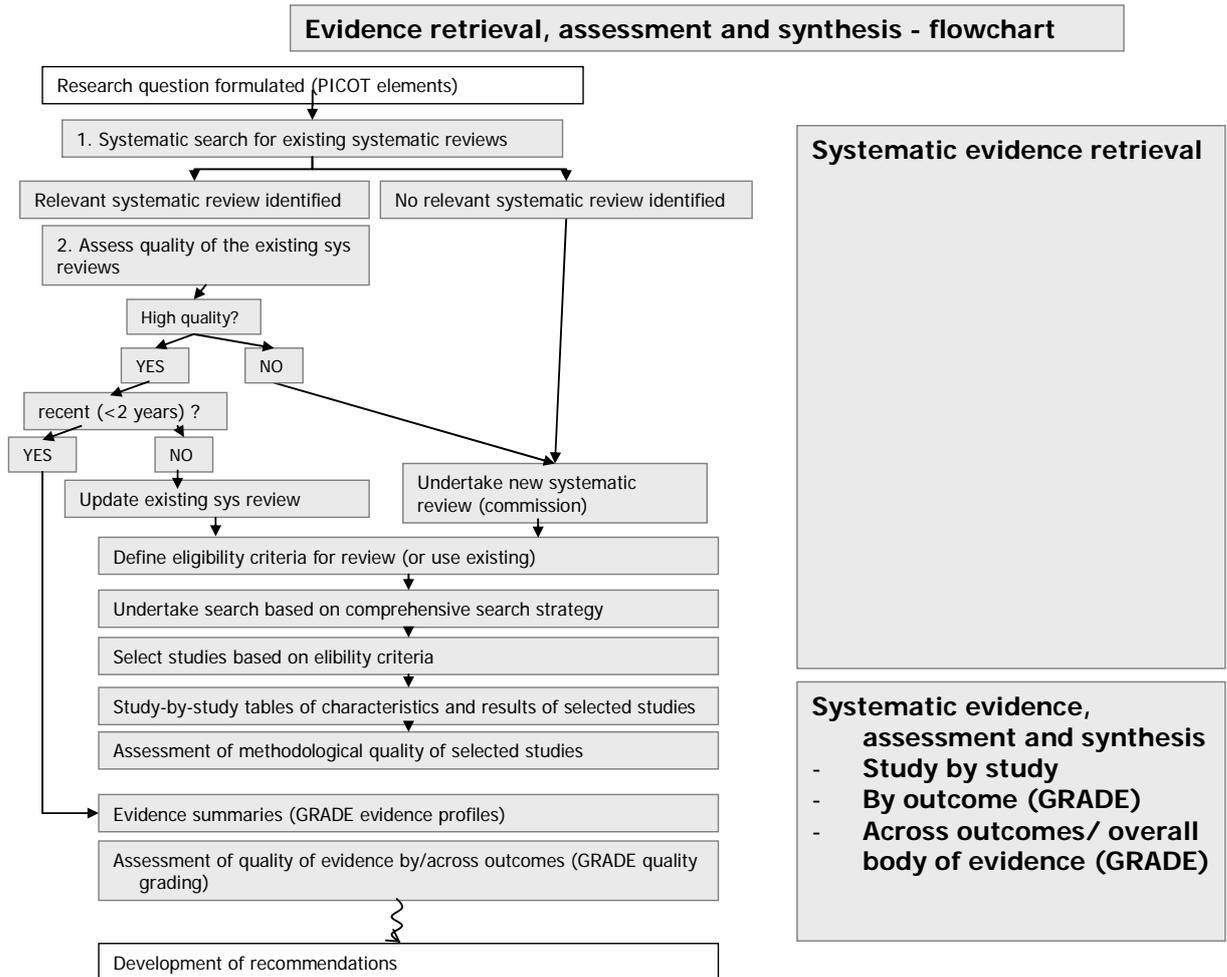
If you are planning an update of an existing WHO guideline, defining the scope should include prioritization of the recommendations that need updating, and those that should be added to or deleted from the previous guideline. Each recommendation you decide to include in your update should be updated according to current WHO guideline development standards outlined in this handbook, including a systematic evidence retrieval process. This can include an update of the systematic review on which the existing recommendation is based, or the selection or production of a new systematic review. Figure 5 gives an overview of the evidence retrieval, assessment and synthesis process.

Additional reading

NICE. *The guidelines manual 2009. Chapter 2. The scope*. London, National Institute for Health and Clinical Excellence, 2009 (available at <http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinenevelopmentmethods/GuidelinesManual2009.jsp?domedia=1&mid=5F482C62-19B9-E0B5-D4282A6200E3200C>).

SIGN 50: a guideline developer's handbook. Section 1. Introduction. Revised edition 2008. Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

Fig. 5. Overview of evidence retrieval, assessment and synthesis



Guideline development group - function and composition

Purpose of the group

It is usually necessary to convene a guideline development group to advise on the content of a WHO guideline. The guideline development group should at a minimum develop and agree on the recommendations and review the completed document. A secondary purpose of a guideline development group is to advocate for implementation of the guideline at country level. The group can meet electronically, but will usually need to have at least one or two face-to-face meetings. The size, composition and function of the group are important considerations in planning the budget.

The functioning of a guideline development group is critical to the quality of a guideline. The role of the group should be:

- to advise on the priority of questions and scope of the guideline;
- to advise on the choice of important outcomes for decision-making
- to comment on the evidence used to inform the guideline;
- to advise on the interpretation of the evidence, with explicit consideration of the overall balance of risks and benefits;
- to formulate recommendations, taking into account diverse values and preferences according to GRADE.

Composition of the group

The guideline development group should be multidisciplinary and should include:

- content experts for all specialties involved;
- methodologists (experts in assessing evidence and developing guidelines; health economists, statisticians as appropriate);
- representatives of potential stakeholders, such as managers and other health professionals involved in the health-care process and who are likely to be end-users of the guideline;
- patients and consumers, although this can be difficult at global level.

For WHO rapid advice guidelines, it is not necessary to convene a formal guideline development group meeting as the time frame for production is too short. However, consultation with external experts – either electronically or face to face – would be ideal.

For standard and full guidelines (including updates and compilations of guidelines), a guideline development group should be convened. It should comprise representatives

of all regions likely to use the guideline. The group should include potential users, content experts, and experts in systematic reviews and guideline development methodology. Inclusion of end-users increases the likelihood of producing a guideline that is appropriate to their needs and thus may contribute to successful implementation.

The selection of the chair of the group is a key decision. People who are experts in the content area of the guideline with strong views about interventions or aspects that will be included should not chair a guideline group. In most situations groups work most effectively if the chair has some knowledge of the content but is particularly expert in facilitating groups and interpreting evidence. Options include having a WHO staff member co-chair the group and ensuring that the chair does not have a veto within the group.

Consumer involvement in WHO guideline development, although challenging at global level, should be encouraged from the start of the guideline process, particularly for guidelines that have a predominantly clinical focus. An increasing number of consumer groups are operating at international level. Many countries have nongovernmental organizations (NGOs) with members who might be available.

The benefits of involving consumers in guideline groups include the ability to formulate questions relevant to consumers, ensuring that relevant aspects of the experience of illness are covered in planning effective treatment (e.g. communicating evidence-based choices to patients and providing information on self-care). Other potential benefits include identifying and prioritizing outcomes of importance, and preparing a consumer-accessible version of the guidelines. Barriers to consumer participation include the lack of suitable consumer groups, time constraints, and the complexity of scientific terminology and other language used by committees. In some cases, such as an illness-related group, illness may preclude patient involvement. This is a factor that has hindered consumer participation.

Experience from organizations such as the United Kingdom's National Institute for Health and Clinical Excellence (NICE) suggests that, for consumers to be active contributors in guidelines groups, there should be more than one consumer in the group and some training beforehand would be helpful. For further information, see the NICE web site at <http://www.nice.org.uk/>. Consumer groups may be funded directly or indirectly by industry so members of patient groups should declare any conflicting interests.

Gender balance should also be considered in the selection of group members.

Inclusion of a methodologist (a person with experience in guideline development processes) is recommended for the planning stages before the guideline development group has been formed. The methodologist can often give valuable advice on group composition. A list of methodologists who have agreed to be contacted by WHO guideline development groups is available at the GRC secretariat

A guideline development group that has been created to update a guideline does not have to include the same persons as the original guideline development group.

It is strongly recommended that one person should be responsible for writing the guideline, with group members commenting on it and suggesting improvements to the text. This will help ensure a standardized approach throughout the process of guideline development. The external peer-review group will also provide comments, as detailed further in this handbook.

The size of a guideline development group should be small enough for effective group interaction but large enough to ensure adequate representation of relevant views. A group of 15–20 is usually feasible and affordable. Additional consultations (outside group meetings) may be held through electronic communication.

Representatives of commercial organizations may not be members of guideline groups. They may be invited to attend parts of the group meetings as observers but must not be present when recommendations are being formulated.

For advice on how to run guideline development group meetings effectively, please see [Appendix I](#).

Additional reading

Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health research policy and systems* 2006;4:15 (doi: 10.1186/1478-4505-4-15).

Hutchings A, Raine R, Sanderson C, Black N. A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services Research and Policy* 2006;11:218–24.

NICE. *The guidelines manual 2007. Chapter 9. Making group decisions and reaching consensus*. London, National Institute for Health and Clinical Excellence, 2007 pp 20–29, 56–59 (available at http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinenevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual_all_chapters.jsp).

SIGN 50: a guideline developer's handbook. Section 5. The guideline development group. Revised edition 2008. Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

Declaration and management of conflicts of interests (COI)

According to the rules in the WHO Basic Documents, all experts participating in WHO meetings must declare any interest relevant to the meeting prior to their participation. The Declaration of interest form to be completed include:

- Personal and non personal (family) financial interest,
- academic interests
- public statements and other activities that may be relevant to the subject of the meeting or guideline.

In addition, anyone invited to participate **in a substantive way** in the development of a guideline must also complete a Declaration of Interests form, and must agree to the publication of the declaration in the guideline. Preparation of systematic reviews and evidence profiles, or contributing to the formulation of recommendations and writing the guideline are considered substantial contributions.

For further information about WHO requirements see: [Office of the Legal Counsel](#)

The current [WHO form for declaration of interests](#) is available at the GRC website.

The DOI forms must be completed prior to any guideline meetings. Following the completion of DOI forms, the declarations are assessed and disclosed in three steps.

- 1) WHO staff need to assess the declared interests *prior to the person participating in the meeting* to determine whether a conflict exists that may preclude or limit the participation of the person in the guideline group.
- 2) Open declaration of interests at the guideline group meeting, so that the group can be aware of interest that exist among the members.
- 3) Reporting of a summary of all declared interests in the final guideline document.

1. Assessment prior to final selection of group members.

The conflict of interest statements of any potential guideline group member should be reviewed initially by the responsible technical officer and the relevant department before finalization of the group composition and invitation to attend any guideline group meeting.

No one may contribute to the development of a guideline until the Declarations of Interest have been reviewed by the WHO guideline steering group, and, if necessary, the Legal Counsel.

The majority of guideline group members should have no conflicts of interest. The chair should have no COI.

The first question is whether any *declared* interests constitute a conflict of interest. What constitutes a potentially significant conflict of interest is a matter of judgment.

Some examples of interests that are clearly a conflict, and that should preclude participation in developing recommendations:.

- owning shares in a company that manufactures a product or technology that may be recommended for use in the guideline (note that there is a financial threshold specified in the reporting form)
- holding a patent on a product or technology that may be recommended for use in the guideline
- having a family member who works for a company that manufactures a product or technology that may be recommended for use in the guideline
- current or past involvement in a major academic program of work that concerns a product or technology likely to be considered in a recommendation, including conducting trials or systematic reviews that recommend a particular product or technology.
- receiving funding from; being or have recently been employed by; consulting for; or acting as an advisor, paid speaker, or opinion leader for a company or organization with an interest in a specific product related to the guideline. This involves receiving any support for travel, professional training or similar.

If a person declares interests that are relevant to the meeting, it is up to WHO - not the guideline development group - to decide whether they can participate in the guideline development.

Based on recent examples during 2008- 2009, we recommend the following -

Participants **should not participate at all** if:

- They declare multiple personal financial links to a single commercial company who owns a technology that is a subject of the guideline
- They declare significant personal financial interests in a single company (>\$10,000) if the company has a technology that is a subject of the guideline

Participants **can participate in the discussion, but not recommendation development** if:

- They declare multiple links with companies with technologies that are the subject of a guideline
- They declare research funding from companies related to the technologies that are the subject of a guideline
- They declare expression of strong public statements, either as academic publications or others, on technologies that are the subject of the guideline

A person with any potential conflict of interest should not chair a guideline group meeting. Participants who are involved in either primary research or conducting a systematic review/s relating to the recommendations in question, should declare so and should not participate in the recommendation ratification.

All such decisions need to be documented *prior to the meeting* and if the Department requires further advice on individual participants, the Legal Counsel should be consulted.

Legal Counsel may advise that:

- the conflict of interest is considered insignificant but must be reported in the final guideline.
- the conflict of interest is significant but related to only some areas of the guideline development group's work. In this case the participant cannot participate when the group considers these areas, and will not have access to the relevant documents.
- the conflict of interest is such as to preclude participation.
- participation in the discussions but not in the ratification of recommendations

In all cases Legal Counsel's advice will be determinative.

All actions taken should be clearly documented in the guideline.

2. Declaration

All Declarations of Interest by the participants of a guideline development group meeting should be provided to all other participants. At each meeting of the group, all members need to provide a verbal summary of their written declaration of interest. We recommend that early on in the group's meeting agenda, before any significant discussion takes place, the Chair should project all the Declarations for the group to review and revise as necessary.

The WHO staff (not the group) will then need to make a judgment as to whether the declared interests are of potential importance with respect to likely recommendations and if so, how to manage the declared conflicts. The decisions made should be clearly documented.

There are no absolute rules about how best to manage conflicts of interest in decision making. Options include excluding the person from participating in the discussion of a particular item, excluding the person from the part of the meeting that formulates the recommendation, or taking no action at all.

Some examples of approaches used in recent WHO meetings are:

Dr N.C. reported being an investigator on trials for GlaxoSmithKline, Quintiles, Uriach and Biomarin but not for any products or products related to those being

considered at the meeting, and also holding shares in Biota. She therefore was excluded would be better from discussion of the late item on antivirals.

Dr M.R. reported having been a consultant for Roche on drug research and development. He is currently a member of a data safety and monitoring board for Roche; receives royalties through the US National Institutes of Health from the use of gossypol for cancer; and is a consultant to several start-up companies, none of which have products on the market. As there were no products related to any of these items on the agenda, no action was required.

Dr A.F. reported having a family member who is an employee of Merck, Sharpe and Dohme, Brazil. He therefore excluded himself from review or discussion of the product applications from Merck on this agenda.

3. Reporting

A summary of how conflicts of interest declarations were collected, any declared conflicts and a brief description of how they were managed must be included in the actual guideline document. If no conflict was declared, this information needs to be provided as well. The GRC will not clear a guideline document that does not contain this information.

Declared conflicts of interest should be reported in the guideline according to the following examples:

Four experts declared an interest in the subject matter of this meeting.

Dr Sanders: his wife is employed by a company that manufactures veterinary products.

Dr Issack: he has limited [define range of \$ value] shareholdings in two pharmaceutical companies that produce antibiotics.

Dr Acarshe has acted as a paid consultant for a producer of veterinary products and for a fast-food chain.

Dr Bywater: he holds a retirement pension and shares in, and conducts occasional consultancies for, two pharmaceutical companies that produce antibiotics, and occasionally consults for a company producing veterinary products. Dr Bywater did not participate in the final day of this workshop during which the final recommendations were discussed and the report was adopted.

The text in which potentially significant conflicts of interest are reported must be approved by Legal Counsel before final review by the GRC.

What to do when you have too many conflicts? Can participants with conflicts attend as 'observers'?

WHO has traditionally relied on 'experts' in its recommendation development, on the basis that there expertise is needed and that the advice will still be objective. Unfortunately, there are many examples where this assumption is simply not true. We therefore recommend that if you need to have input from experts who have conflicts or who are not objective, you use an 'evidence jury'. To do this you should make sure that at least two thirds of your group have no conflicts of interest. They become your jury. The experts then can present their views and evidence to the jury who eventually develops the recommendations - but without the experts being present for that part of the meeting and discussion. Logistically it means that you divide your meeting into (at least) two parts - the first with everyone present and the second (closed) session, for recommendation development and ratification, that only the jury attends.

You can extend this to a three part meeting if necessary, to include a session - preferably on the first day - where commercial organisations with interest in the guidelines might be offered an opportunity to present information to the meeting. We strongly recommend that pharmaceutical company representatives should not attend any other part of a guideline panel meeting, as observers or otherwise.

Additional reading

Bekelman JE, Li Y, Gross C. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA* 2003;289:454-65. <http://jama.ama-assn.org/cgi/reprint/289/4/454>.

Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002;287:612-17. <http://jama.ama-assn.org/cgi/reprint/287/5/612>.

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70. <http://www.bmj.com/cgi/reprint/326/7400/1167>.

Sniderman AD, Furberg CD. [Why guideline-making requires reform](http://jama.ama-assn.org/cgi/reprint/301/4/429). *JAMA* 2009;301:429-31. <http://jama.ama-assn.org/cgi/reprint/301/4/429>.

Steinbrook R. Controlling conflict of interest--proposals from the Institute of Medicine. *NEJM* 2009;360:2160-63. <http://content.nejm.org/cgi/content/full/360/21/2160>.

Formulating questions and choosing outcomes

The selection of questions (and the components of the questions) that are to be addressed in the guideline has major consequences for the scope of the guideline. The questions will drive the direction (inclusion and exclusion of data) and determine the type of information that will be searched for and that will be assessed. The questions are also the starting point for formulating the recommendations. It is very important that the questions are clear and well defined, and that there is group agreement about them.

Updating a guideline may include a change of scope; not only questions but also the selection of critical outcomes may differ from the original guideline.

It is helpful to start by dividing the types of information and questions into three main categories:

Definition/background questions

e.g. What is human papilloma virus (HPV) infection?

Facts/foreground questions

e.g. What is the effectiveness of an HPV vaccine?

Recommendation/decision

e.g. Should we use HPV vaccine?

Guidelines usually include all three categories of question.

The questions to be covered by the guideline should be identified on the basis of clinical or policy needs and input from clinicians and other experts, including programme managers and partner agencies. Input from consumer or patient groups may also be helpful. Generally, questions should focus on areas of controversy that need to be answered by the guideline or on areas where changes in policy or practice are needed. To develop the questions for the guideline, the steering group should:

- specify the purpose of the guideline (it may be helpful to divide the objectives into general and specific objectives);
- draft the foreground and background questions as researchable questions;
- prioritize the questions;
- determine which questions will need systematic reviews.

The number of questions that need systematic reviews will be a major determinant of the time and resources needed to complete the guideline.

The initial list of types of question to be covered will probably be a long one. Some examples could be:

- What are the phenomena associated with the problem? (background)
- What is the frequency of the problem? (background)
- What causes the problem? (etiology)
- Does this person have the problem? (diagnosis)
- What happens if you get the problem? (prognosis)
- How can we treat the problem? (intervention)
- What policies should we introduce to alleviate the problem? (policy intervention)

To turn these into answerable questions, the PICOT framework is useful:

P opulation <i>In women without HPV</i>	(What factors are essential?)
I ndicator/ i ntervention <i>Does HPV vaccine</i>	(Specific intervention or class?)
C omparator <i>Compared with no treatment</i>	(Compared to nothing or standard treatment)
O utcome <i>Reduce rates of carcinoma in situ</i>	(Patient-relevant outcomes, including both benefits and potential side-effects)
T ime <i>For life?</i>	(Short term and long term)

An example of a PICOT question on health policy could be:

In a rural population in low- and middle-income countries (**p**opulation), does paying higher salaries to health workers (**i**ntervention), compared with paying standard salaries (**c**omparator), increase the number of health workers in rural areas (**o**utcome) in the long term (10 years) (**t**ime)?

This format can also be used, with slight modifications, for questions on prevalence and incidence, etiology (exposure-outcome) and diagnosis. For instance:

- *In women in Uganda (P), what is the frequency of breast cancer (O) each year (T)?*
- *In men over 40 years of age (P), what is the rate of lung cancer (Outcome) after 10 years (Time) in smokers (Exposure) versus non-smokers (Comparator)?*
- *In babies born to HIV-positive women (P), does screening with a new rapid diagnostic test (I, C) before the age of six weeks (T) accurately detect disease?*

The type of question will also determine the type of evidence that should be sought. For instance:

- **What are the phenomena/experiences/views that influence patient behaviour?**
 - *Observational studies (e.g. qualitative research)*
- **What is the frequency of the problem? (frequency)**
 - *Random (or consecutive) sample*
- **What causes the problem? (etiology)**
 - *Observational studies*
- **Does this person have the problem? (diagnosis)**
 - *Random (or consecutive) sample with gold standard*
- **Who will get the problem? (prognosis)**
 - *Follow-up of inception cohort*
- **How can we alleviate the problem? (intervention/therapy)**
 - *Randomized controlled trial(s) or systematic reviews of such trials?*
- **How can we alleviate the problem? (intervention/policy)**
 - *Randomized controlled trial(s) and/or non-randomized studies*

Other possible guideline questions are:

- **How is the effective treatment best delivered? (organizational intervention)**
 - *Randomized controlled trials or systematic reviews of such trials*
- **How are the pros and cons of treatment best communicated with the patient or family member? (communication intervention or strategy)**
 - *Randomized controlled trials or systematic review of such trials*

The facts/foreground questions are the most important ones for a guideline. They are used to inform the recommendation/decision and they will require a systematic review and quality assessment of the evidence using the GRADE approach.

Information (including systematic reviews) about the background questions can inform adaptation issues, values and preferences, clinical needs and baseline risks.

After the guideline group has reached consensus, the draft questions should be sent to an international panel of clinicians, researchers, programme managers and consumers (the extended external group) for comment and revision. These reviewers should be asked to identify any omissions from the key questions that should be covered by the guideline, or relevant outcomes to be considered in developing recommendations.

Additional reading

NZGG. Question formulation for clinical practice guidelines. In: *Handbook for the preparation of explicit evidence-based clinical practice guidelines*. Wellington, New Zealand Guidelines Group, 2001, pp 15–21 (available at http://www.nzgg.org.nz/download/files/nzgg_guideline_handbook.pdf).

Kaifeng AK. Principles of evidence based medicine. *Arch Dis Child* 2005;90:837–840.

O'Connor D, Green S, Higgins JPT, eds. Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, eds. *Cochrane handbook of systematic reviews of intervention*. Version 5.0.1 (updated September 2008). The Cochrane Collaboration, 2008 (available at www.cochrane-handbook.org).

Choosing and rating outcomes

In addition to reviewing the questions, the external review group should identify the key outcomes that need to be considered in making the recommendations. The purpose of this is to identify the outcomes that will be critical for making decisions and recommendations, and to identify the data that should be sought through evidence retrieval and synthesis.

It is important to focus on the outcomes that are important to patients, and to avoid the temptation to focus on those outcomes that are easy to measure and often reported (unless those are also the important outcomes).

The guideline steering group should make an initial list of possibly relevant outcomes, including both desirable and undesirable effects. One method of facilitating the discussion of which outcomes are the critical ones is to ask the group members to score the relative importance of each outcome from 1–9, where 7–9 indicates that the outcome is critical for a decision, 4–6 indicates that it is important and 1–3 indicates that it is not important. The average score for each outcome can be used to determine the relative importance of each outcome, although it is helpful to provide the range of results as well. Sometimes people with different perspectives (patients, physicians, researchers, policy-makers, pharmaceutical companies) have different opinions about which outcomes are important. Therefore all the groups should be present during the discussion on the selection of critical outcomes. The guideline development group should be asked to identify additional important outcomes that have not been included in the list of potential outcomes previously identified by the guideline steering group.

It can be useful to have an early hearing of the important stakeholders regarding the selection of the critical outcomes.

If necessary, the final rating of outcomes can be reviewed and confirmed at a later stage (e.g. at a guideline development group meeting).

Additional reading

Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Ascham J et al. Consensus development methods and their use in clinical guideline development. *Health Technology Assessment* 1998;2(3) (available at <http://www.nchta.org/fullmono/mon203.pdf>).

Schünemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development. 6: Determining which outcomes are important. *Health Research Policy and Systems* 2006;4:18 (doi:10.1186/1478-4505-4-18).

Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM et al. (WHO Rapid Advice Guideline Panel on Avian Influenza). WHO rapid advice guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. *Lancet Infectious Diseases* 2007;7:21–31.

WHO recommendations for the prevention of postpartum haemorrhage. Geneva, World Health Organization, 2007, pp 6–8 (available at [http://whqlibdoc.who.int/hq/2007/WHO MPS 07.06 eng.pdf](http://whqlibdoc.who.int/hq/2007/WHO_MPS_07.06_eng.pdf)).

Evidence retrieval, assessment and synthesis

A summary of all relevant research evidence is essential when developing a recommendation. The most important type of evidence is that on the effect of interventions being considered in the recommendation.

The summary of this research evidence should be in the form of a systematic review. In contrast to narrative reviews, systematic reviews address a specific question and apply a rigorous scientific approach to the selection, appraisal and synthesis of relevant studies.

A systematic review requires a protocol that describes:

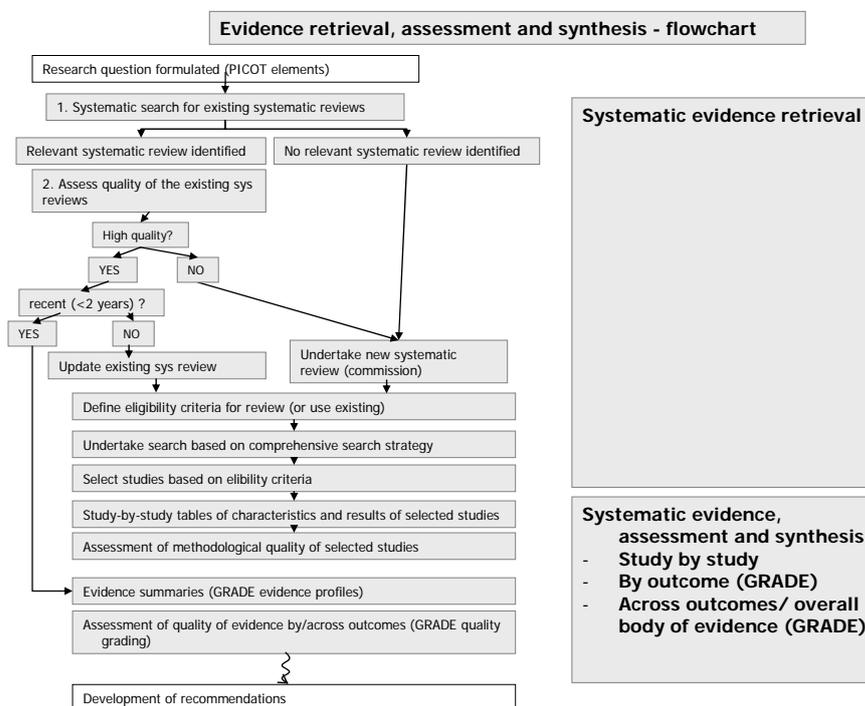
- the search strategy used to identify all relevant published – and unpublished – studies;
- the eligibility criteria for the selection of studies;
- how studies will be critically appraised for quality;
- an explicit method of synthesis of results and, if feasible, a quantitative synthesis of the results of studies to estimate the overall effect of an intervention (meta-analysis).

Systematic reviews, if conducted properly, reduce the risk of selective citation and improve the reliability and accuracy of decisions.

In order to help with the [critical appraisal of systematic reviews](#), please refer to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria for reporting of systematic reviews (Appendix III). You may also wish to visit the web site of [PRISMA](#) to access the [checklist](#).

The process of evidence retrieval, assessment and synthesis is described in further detail below and is summarized in Figure 6.

Fig. 6. Overview of the evidence retrieval, assessment and synthesis process



Prioritizing evidence retrieval

Trying to retrieve evidence to support every recommendation in a guideline may simply not be feasible. This is where it becomes important to identify priority questions or issues that the guideline should address (see section on [defining the scope](#)).

For WHO guidelines, controversial questions – or those where there is a need to change practice – should be supported by evidence summaries prepared according to the GRADE approach and based on systematic reviews. It is less critical to provide a full review (e.g. for a section on the epidemiology or pathology of a disease) if this is for background information only.

Evidence retrieval

Existing systematic reviews

Existing reviews should be used wherever possible and should be updated if necessary. The first step in evidence retrieval is to identify relevant systematic reviews for each of your questions. Before starting a major process of evidence retrieval for a WHO guideline, it is suggested that you start with a search for existing guidelines. If they are of reasonable quality, they should have lists of references, including systematic reviews, which can be used to assist in the search.

The search for existing systematic reviews can be done in-house, or can be subcontracted to a group preparing the evidence summaries. Systematic reviews can include all types of study design. Consider which types are most relevant to your questions.

The most readily accessible biomedical database is PubMed. The PubMed “Clinical Queries” or “Special Queries” options permit specific searches to be set up to identify systematic reviews of different types of studies. This includes searches of the Cochrane Database of Systematic Reviews. Systematic reviews of policy interventions (such as pricing of pharmaceuticals) may be difficult to find, and other search strategies will be needed. The Campbell Collaboration has a database of reviews of effectiveness of social and educational policies and practices. When in doubt, consult a WHO librarian or other expert in information retrieval.

The search strategy should be well documented and should specify:

- the details of the databases (including web sites) to be searched, and the search strategy to be applied to each database;
- the details of each strategy as actually performed, with the date on which the search was conducted and/or updated (this description must be included in the final guideline).

Adequacy of systematic reviews

Once the reviews are retrieved, they should be checked for:

- relevance (to the questions to be addressed in the recommendations);
- timeliness (assessed by date of last update);
- quality (assessed by a standard critical appraisal instrument).

Checklists, such as the one developed by the Oxford Centre for Evidence-Based Medicine (see [Appendix II](#)), should be used for critical appraisal of systematic reviews. The checklist can be accessed at <http://www.cebm.net/index.aspx?o=1157>.

If there are several relevant systematic reviews, use the most recent one of high quality. If the identified systematic reviews are all of low quality, consider commissioning a new one. If the review is of high quality but more than two years old, consider updating the review to include more recent evidence.

Commissioning systematic reviews

A new systematic review is needed if:

- there are key questions to be covered in the guideline that are not answered by an existing systematic review;
- the only relevant and high-quality systematic review is over two years old.

Preparing systematic reviews is time-consuming and requires technical capacity and resources. It is suggested that this should not be done routinely by WHO staff unless there is a specific reason for doing so and time and resources are available. (See Figure 6 for the detailed steps required in undertaking a systematic review.) Systematic reviews used in WHO guidelines must be of high quality and should be guided by the standards outlined by the Cochrane Collaboration in the [Cochrane handbook](#). Following the elements outlined in the systematic review appraisal sheet from the Oxford Centre for Evidence-Based Medicine ([Appendix II](#)) also helps to ensure that the systematic review is of an appropriate standard, with a low risk of producing biased results.

Conducting a systematic literature search

Conducting a systematic literature search is a complex and time-consuming task that should ideally be commissioned externally. If this is not possible, always seek advice from an information retrieval specialist. Important decisions need to be made about the databases to be searched and the search strategy to apply.

Generally, the search strategy is guided by the elements of the research question (population, intervention, comparator, outcomes) and the types of studies you are looking for. For a WHO guideline, it is important to search for studies from developing countries as well as from more standard literature sources. Some journals are not well represented in PubMed and commercial databases such as EMBASE and CAB Abstracts. Regional databases grouped under the general heading of the Global Health Index <http://www.who.int/ghl/medicus/en/> contain unique citations and full-text articles. WHO's regional offices have supported the development of these indices to highlight the health research of developing countries. Most journals indexed by regional databases are not indexed in PubMed or other databases.

Retrieval of “grey literature” – such as Ministry of Health reports, case studies and unpublished studies – is best based on the results of searches for journal articles from PubMed, regional databases and other sources. Author and institutional names can be used in internet search engines to identify grey literature. It is also important to scan key web sites individually as general search engines do not retrieve all the relevant information on a web site. It is important to obtain training in using search engines to be able to limit research results to pertinent information. The combination of efficient use of search engines and targeted web sites and authors is much more effective for identifying unique information than large unfocused searches. Personal contact with key experts will help identify sources of information not found in the published journals or cited on web sites.

For search strategies of systematic reviews of public health interventions, see the relevant Cochrane guidance.^b

^b Armstrong R, Waters E, Jackson N, Oliver S, Popay J, Shepherd J et al. *Guidelines for systematic reviews of health promotion and public health interventions. Version 2.* Melbourne University of Melbourne, 2007

In the final guideline document, a summary of the key elements of the systematic review should be provided together with links and references. The summary should include details of the search terms used and the date on which the search was conducted and/or updated.

Evidence retrieval for compilations of recommendations

When a guideline is developed on the basis of existing recommendations, the selection of these recommendations should be done in a systematic and transparent way, similar to the process described above.

The protocol should describe:

- the search strategy used to retrieve existing guidelines;
- how the relevance and quality of guidelines and their underlying evidence will be assessed;
- how the selection of guidelines will be made if there are more than one on a given topic;
- how you will proceed if you find no guideline on a question of relevance to the compilation (a systematic retrieval process for existing reviews or primary studies should be considered, as above).

Evidence assessment

Assessing the evidence that has been retrieved is a crucial step that enables the guideline development group to formulate recommendations. This assessment is based on the systematic review(s) you have done or commissioned. From these review(s), evidence summaries must be prepared for the guideline development group. It is easiest to do this using tables.

If the evidence retrieval has identified systematic reviews as the basis for developing recommendations, evidence summaries (or profiles) should be created using the GRADE approach. This approach allows a structured and transparent assessment of the quality of evidence for each outcome. For each question, data should be extracted from the systematic review for all the outcomes (benefits and harms) that were rated as important. If there is more than one high-quality, relevant and up-to-date systematic review, start with the best one and supplement it as needed with additional data from the systematic reviews that meet the selection criteria determined previously.

Using observational studies

If you are developing recommendations about clinical interventions, systematic reviews of randomized trials are the preferred starting point. However, the experience of most WHO guideline groups so far suggests that randomized trials alone do not provide sufficient information for developing global recommendations. Observational studies may need to be included in the evidence retrieval process, particularly to provide evidence of the feasibility of interventions in different settings and of the experience of using interventions in real-world populations. If you are planning to include observational studies in your evidence synthesis, you are recommended to seek help from an information retrieval specialist since these studies are more difficult to identify systematically than randomized controlled trials.

Assessing the quality of observational studies is described in the [GRADE handbook](#). The key issue to note is that observational studies of interventions have an increased risk of bias and generally overestimate the size of the effect of any intervention. Therefore, relying on observational studies alone to assess whether an intervention works is unwise. However, if the systematic review has identified only a few small

randomized controlled trials, supplementing the estimate of effect size from the those trials with the results of the observational studies may increase a guideline group's confidence in the effect of the intervention and may provide important information about the feasibility of the intervention in resource-poor settings.

Statistical methods for combining results of observational studies are more complex than the methods used for combining randomized controlled trials. It is possible to pool data and if results from observational studies are variable in this format, they can be reported in a standard GRADE evidence profile. If the results are not combined, it is recommended to present the results of observational studies in a study-by-study evidence profile (see the WHO guideline on Pharmacological Management of Influenza, 2009, for examples).

It is recommended to seek advice from the GRC secretariat if you plan to use observational studies in your guideline development.

Evidence profiles/summaries of single studies

If you do not have the results of studies presented as the results of a meta-analysis, evidence profiles can be prepared using tables of single studies. The format can be a GRADE table or the format used in study tables or summary-of-findings tables for Cochrane systematic reviews. For guideline groups, we suggest a GRADE table or summary-of-findings table format as it is easier to understand.

For summaries of studies of the epidemiology of a condition, or prognostic studies, use a study-by-study table (systematic review table with data extracted by study).

Draft evidence summaries should be sent to the members of the guideline development group before the meeting. Group members should be asked to identify any relevant evidence that is missing from the summaries.

Additional reading

Akobeng AK. Understanding randomized controlled trials. *Archives of Disease in Childhood* 2005;90:840–844.

Akobeng AK. Understanding systematic reviews and meta-analysis. *Archives of Disease in Childhood* 2005;90:845–848.

Centre for Reviews and Dissemination. *Systematic reviews. Chapter 3. Public health interventions* (available at <http://www.york.ac.uk/inst/crd/SysRev/SSL/WebHelp/SysRev3.htm>).

Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development. 7: Deciding what evidence to include. *Health Research Policy and Systems* 2006;4:19 (doi:10.1186/1478-4505-4-19) (available at <http://www.health-policy-systems.com/content/4/1/19>).

Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development. 8: Synthesis and presentation of evidence. *Health Research Policy and Systems* 2006;4:20 (doi:10.1186/1478-4505-4-20) (available at <http://www.health-policy-systems.com/content/4/1/20>).

Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development. 9: Grading evidence and recommendations. *Health Research Policy and Systems* 2006;4:21 (doi:10.1186/1478-4505-4-21) (<http://www.health-policy-systems.com/content/4/1/21>).

NZGG. Finding the evidence – literature searching for guideline development. In: *Handbook for the preparation of explicit evidence-based clinical practice guidelines*. Wellington, New Zealand Guidelines Group, 2001, pp 23–37 (available at http://www.nzgg.org.nz/download/files/nzgg_guideline_handbook.pdf).

Deurenberg R, Vlayen J, Guillo S, Oliver TK, Fervers B, Burgers J, on behalf of the SEARCH group. Standardization of search methods for guideline development: an international survey of evidence-based guideline development groups. *Health Information and Libraries Journal* 2008;25:23–30.

NHMRC. *How to use the evidence: assessment and application of scientific evidence*. Handbook series on preparing clinical practice guidelines. Canberra, National Health and Medical Research Council, 2000 (available at <http://www.nhmrc.gov.au/publications/synopses/files/cp69.pdf>).

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MB et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *Journal of the American Medical Association* 2001;286:821–30.

NZGG. Assessing and applying the evidence. In: *Handbook for the preparation of explicit evidence-based clinical practice guidelines*. Wellington, New Zealand Guidelines Group, 2001, pp 42–54 (available at http://www.nzgg.org.nz/download/files/nzgg_guideline_handbook.pdf).

Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G et al. The GATE frame: critical appraisal with pictures. *Evidence-based Medicine* 2006 ;11:35–38.

Sawaya GF, Guirguis-Blake J, LeFevre M et al. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Annals of Internal Medicine* 2007;147:871–875.

Haute Autorité de Santé. *Guide méthodologique. Méthodes quantitatives pour évaluer les interventions visant à améliorer les pratiques*. Saint-Denis La Plaine, Haute Autorité de Santé, 2007 (available at http://www.has-sante.fr/portail/upload/docs/application/pdf/eval_interventions_ameliorer_pratiques_guide.pdf).

Higgins JPT, Altman DG, eds. *Cochrane handbook for systematic reviews of interventions. 8: Assessing risk of bias in included studies*. Preliminary draft, 3 December 2007 (available at http://www.cochrane.org/resources/handbook/Handbook5_Bias_V10.pdf).

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *British Medical Journal* 2008;336:1106–1110.

Grading the quality of evidence and the strength of recommendations

WHO uses the GRADE approach for developing recommendations. The software (GRADEprofiler) to facilitate the process is freely available and can be downloaded from <http://www.ccs-ims.net/revman/gradepro/gradepro>. Further information can be obtained by emailing mail@gradeworkinggroup.org.

It is recommended to download GRADEprofiler and to open the [GRADEprofiler help file](#). There you will find:

- the GRADE handbook for grading the quality of evidence and the strength of recommendations;
- a guide to getting started;
- information on creating GRADE evidence profiles;
- information on creating tables for summaries of findings;
- how to manage evidence profiles;
- an overview of the GRADE approach;
- glossary, references, technical requirements, support, and details of the GRADE working group.

This chapter refers to different sections of the GRADE handbook (Schünemann H, Brożek J, Oxman A, eds. [GRADE handbook for grading quality of evidence and strength of recommendation. 3.2](#) [updated March 2009]).

You are encouraged to contact the GRC secretariat at the planning stage to enquire about planned GRADE workshops at WHO or GRADE experts who are available to support you in this activity. In general, it is recommended that a person in the team who is familiar with the GRADE approach should take the lead in this process. The GRC secretariat offers GRADE workshops throughout the year. Please contact the GRC secretariat for more information at grcinfo@who.int.

Fig. 7. Overview of the GRADE process

(see "[Overview of the GRADE approach](#)" in: Schünemann H, Brożek J, Oxman A, eds. *GRADE handbook for grading quality of evidence and strength of recommendation. 3.2* [updated March 2009].)

Evidence assessment and synthesis, and formulation of recommendations:
The GRADE approach



The GRADE approach has two main steps – the evaluation of the [quality of evidence](#) and the [summary of findings](#).

Evaluating the quality of evidence

Quality of evidence is defined as the “extent to which one can be confident that an estimate of effect or association is correct”. The following criteria are taken into account when assessing the quality of evidence:

- the study design (and any limitations of studies in terms of their conduct and analysis);
- the consistency of the results across the available studies;
- the precision of the results (wide or narrow confidence intervals);
- the directness (or applicability or external validity) of the evidence with respect to populations, interventions and settings where the proposed intervention may be used;
- the likelihood of publication bias.

For observational studies there are additional criteria, namely:

- the magnitude of the effect;
- the presence or absence of a dose-response gradient;
- the direction of plausible biases.

Quality of evidence is categorized as high, moderate, low or very low (see Table 1 for definitions of these grades).

Table 1. Grades of the quality of evidence

Grade	Definition
High	Further research is very unlikely to change confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on confidence in the estimate of effect and is likely to change the estimate.
Very low	Any estimate of effect is very uncertain.

Summarizing findings

The second half of a GRADE table consists of the summary of findings – the results of the studies (i.e. the pooled estimates), using both relative and absolute measures. For each outcome, you will need to extract the results from the review(s).

The following information is needed for dichotomous outcomes:

- the total number of patients in each group;
- the total number with event;
- an estimate of the control group risk (control event rate);
- effect size (relative risks or odds ratios, absolute differences and 95% confidence intervals).

Wherever possible, use the results as presented in the review rather than recalculating, assuming that the review methods are correct. Nevertheless, you may have to calculate absolute effect sizes and numbers needed to treat if you wish to represent these to the guideline group.

For continuous outcomes you will need:

- the total number of patients in each group;
- a summary estimate of effect (weighted mean difference or standardized mean difference) and 95% confidence interval.

(See “[Creating GRADE evidence profiles - summarizing the evidence](#)” in: Schünemann H, Brożek J, Oxman A, eds. *GRADE handbook for grading quality of evidence and strength of recommendation*. 3.2 [updated March 2009].)

GRADE profiler will create tables that show the results of your quality assessment and summary of findings. It is advisable for one reviewer to extract data from the systematic reviews or from single studies and to prepare drafts of the GRADE evidence profiles with detailed footnotes explaining the judgments that were made. Each judgment should be made explicit and should be available to the reader in order to increase the transparency of the process. These footnotes should be checked by at least one other member of the team.

Using GRADE tables

The quality of the evidence is assessed for each outcome that has been identified and rated previously by the group. To facilitate this process, the GRADE system provides templates for tables.

(See “[Creating GRADE evidence profiles](#)” in: Schünemann H, Brożek J, Oxman A, eds. *GRADE handbook for grading quality of evidence and strength of recommendation*. 3.2 [updated March 2009].)

To complete a GRADE table for quality assessment, the following should be considered:

- the study design;
- the limitations of the studies in terms of their conduct and analysis;
- the consistency of the results across the available studies;
- the directness (or applicability or external validity) of the evidence with respect to the populations, interventions and settings where the proposed intervention may be used;
- the precision of the summary estimate of effect;
- other considerations (see below).

Below are the aspects that should be considered when [evaluating the quality of evidence](#). When completing GRADE tables these elements will have to be assessed for each outcome separately.

Study [design](#)

Studies are broadly classified into two types:

- randomized controlled trials;
- observational studies, including interrupted time-series (or quasi-experimental design), cohort studies and case-control studies, and other types of design such as case series and case reports.

The design is the baseline for rating the quality of evidence. If you have studies of more than one design reporting the outcome, you should have a separate row in your table for each type.

Evidence based on randomized controlled trials begins as high quality and evidence from observational studies begins as low quality.

Study [limitations](#)

Having defined the type of studies, one should then consider how well they were performed and analyzed.

For randomized controlled trials, the main criteria for assessing trial limitations are:

- whether concealment of allocation to treatment group is adequate;
- whether participants and investigators were blinded, especially if the outcomes are measured subjectively and subject to bias;
- whether an intention-to-treat analysis is reported;
- whether all withdrawals and patients lost to follow-up are accounted for;

- whether the trial was stopped early for benefit.

For observational studies, the main criteria depend on the design (i.e. case-control or cohort studies). For both designs, the methods used to select the population in the study and the comparability of the two groups are important. For case-control studies, the method of determining exposure to the factor of interest also needs to be evaluated. For cohort studies the method of measuring outcomes should be evaluated.

For studies of diagnostic accuracy, the [QUADAS tool](#)^c can be used.

Appraisal tools for systematic reviews, randomized controlled trials, and observational and diagnostic studies can be found at <http://www.cebm.net/index.aspx?o=1157>.

Although rating limitations can be seen as a continuum, within GRADEprofiler you need to decide on the level, which requires a degree of judgment. For instance:

- “No limitations” generally means that the majority of studies meet all the minimum quality criteria for the design. The implication of this is that the rating of quality of evidence remains the same as the initial assessment.
- “Minor limitations” applies when minor flaws are found when analysing how the available studies were designed and performed. For example, allocation concealment may not be reported for one study out of several in a systematic review, or a study could be non-blinded but nevertheless reports objective outcomes. If you decide there are minor limitations, these should be noted in a footnote but they would not usually downgrade the quality.
- “Serious limitations” means that one of the minimum criteria for quality is not met by the majority of studies in the review. This results in a -1 score for the overall quality rating (e.g. “high” becomes “moderate”).
- “Very serious limitations” means that at least two of the criteria proposed as potential study limitations are present in the majority of studies in the review. This results in a -2 score for quality.

The criteria that are used for downgrading the quality of evidence and the reason for the assessment should be explained in a footnote to the table.

Assessing [consistency](#)

Consistency relates to whether the results are similar across studies. Differences in the direction of effect, the size of the differences in effect, and the significance of the differences guide the (inevitably somewhat arbitrary) decision about whether important inconsistency exists. If all the results of the studies for one outcome are in the same direction with overlapping confidence intervals, there is unlikely to be significant inconsistency. To evaluate the degree of consistency of the results of the available studies, the direction and size of the effect for each outcome should be evaluated. If a formal meta-analysis was conducted, the result

^c Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3:25 (available at <http://www.biomedcentral.com/1471-2288/3/25>).

of the test for heterogeneity can be used to help assess consistency. Variability or inconsistency in results may arise from differences in the populations in the studies, in the interventions or in outcomes.

If there is inconsistency in the results, such as the largest trial showing results that contradict smaller trials, then a -1 score should be applied. If the results are very heterogeneous, “very serious” should be chosen, which will downgrade the evidence for this outcome by two levels.

If only one study is present, consistency is not applicable as a criterion.

Assessing directness

Directness, generalizability, external validity of study results and applicability are all synonymous. There are two types of indirectness:

- Indirect comparison occurs when a comparison of intervention A versus B is not available, but A was compared with C and B was also compared with C. Such trials allow indirect comparisons of the magnitude of effect of A versus B. Such evidence is of lower quality than head-to-head comparisons of A and B would provide.
- Indirect population, intervention, comparator or outcome arise when the question being addressed by the guideline development group or by the authors of a systematic review is different from the available evidence regarding the population, intervention, comparator or outcome.

To determine whether important uncertainty exists, consider whether there is a compelling reason to expect important differences in the size of the effect. Because many interventions have more or less the same relative effects across most patient groups, criteria and judgements of directness should not be excessively stringent.

For some therapies (e.g. behavioural interventions in which cultural differences are likely to be important) directness is more likely to be a problem. Similarly, reviewers may identify uncertainty about the directness of evidence for drugs that differ from those in the studies but that are within the same class. Studies using surrogate outcomes generally provide less direct evidence than those using outcomes that are important to people. It is therefore prudent to use much more stringent criteria when considering the directness of evidence for surrogate outcomes.

For WHO guidelines, directness is a very important dimension that has relevance for the implementation of study results in actual practice. The judgement about whether there is “some uncertainty” or “major uncertainty” about directness can be challenging. Although there are no firm guidelines, if there is only one study (e.g. in a developed-country setting and the intervention is likely to be altered according to setting) this would be rated as “major uncertainty” (and therefore scored as -2).

Imprecision

Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect. In this case the quality of the evidence is lower than it otherwise would be because of uncertainty in the results.

When event rates are very low, 95% confidence intervals around relative effects can be very wide, but 95% confidence intervals around absolute effects may be narrow. In the latter case, one should not downgrade the quality of evidence for imprecision.

Please see the GRADE handbook for details on how to evaluate imprecision.

Other considerations

Publication and reporting bias

Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies or selective reporting of outcomes.

Reporting bias arises when investigators fail to report studies they have undertaken (typically those that show no effect) or neglect to report outcomes that they have measured (typically those for which they observed no effect). Methods to detect the possibility of publication bias in systematic reviews exist, although authors of the reviews and guideline panels must often guess about the likelihood of reporting bias.

A situation that should arouse suspicion of reporting bias is when published evidence is limited to a small number of trials, all of which were funded by a for-profit organization. In such a situation, consider the extent to which evidence about the magnitude of the effect is uncertain due to selective publication of studies or reporting of outcomes. If this is likely, downgrade the quality rating by one or even two levels.

Large effects

When methodologically strong observational studies yield large or very large and consistent estimates of the magnitude of a treatment or exposure effect, we may have confidence in the results. In such situations, the weak study design is unlikely to explain all the apparent benefit or harm, even though observational studies are likely to provide an overestimate of the true effect.

The larger the magnitude of effect, the stronger the evidence becomes. Only studies with no threats to validity (not downgraded for any reason) can be upgraded.

Dose-response curve

The presence of a dose-response gradient may increase confidence in the findings of observational studies and thereby increase the quality of evidence. However, this applies only to studies that are not downgraded for any reason. To rate the presence of a dose-response gradient:

- if there is no evidence of dose-response gradient, there is no change;
- if there is evidence of a dose-response gradient, upgrade the evidence for this outcome by 1 level.

Direction of confounding factors

On occasion, all plausible biases from observational studies may tend to underestimate the true treatment effect. For instance, if only sicker patients receive an experimental intervention or exposure, yet they still improve, it is likely that the actual intervention or exposure effect is larger than the data suggest.

Only studies with no threats to validity (not downgraded for any reason) can be upgraded. To rate the effect of all plausible residual confounding:

- if there is no evidence that the influence of all plausible residual confounding would reduce the observed effect, there is no change;

- if there is evidence that the influence of all plausible residual confounding would reduce the observed effect, upgrade the evidence for this outcome by 1 level.

Table 2 summarizes the rating criteria. The final score is determined by adding the additional ratings to the original baseline score.

Table 2. GRADE quality of evidence: assessment criteria

Quality of evidence	Study design	Lower the quality in presence of	Raise the quality in presence of
High	Randomized trial	Study limitations: -1 Serious limitations -2 Very serious limitations	Strong association: +1 Strong, no plausible confounders, consistent and direct evidence
Moderate			
Low	Observational study	-1 Important inconsistency Directness:	+2 Very strong, no major threats to validity and direct evidence*** +1 Evidence of a dose-response gradient
Very low	Any other evidence	-1 Some uncertainty -2 Major uncertainty -1 Imprecise data -1 High probability of reporting bias	

Assessing evidence for other types of questions

In the case of diagnostic test questions, an approach for using GRADE to assess the evidence for diagnostic tests has been developed.^d

Assessing the quality of observational studies for questions of prognosis and risk factors can be done using the criteria for observational studies described in the section on “Study limitations”.

Evidence such as countries case studies should only be described, and not assessed for quality.

Additional reading

Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA et al. ATS Documents Development and Implementation Committee. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *American Journal of Respiratory and Critical Care Medicine* 2006;174:605–614.

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE et al. for the GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *British Medical Journal* 2008;336:1106–1110.

Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schünemann H. An emerging consensus on grading recommendations? *Evidence-based Medicine* 2006;11:2–4.

Schünemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R et al. Transparent development of the WHO rapid advice guidelines. *PLoS Med* 2007;4:e119.

WHO Recommendations for the prevention of postpartum haemorrhage. Geneva, World Health Organization, 2007 (Annex 4) (available at http://www.who.int/making_pregnancy_safer/publications/PPHTables2.pdf).

^d Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, Muti P et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evidence-based Medicine* 2008;13:162–163 (available at <http://ebm.bmj.com/cgi/reprint/13/6/162-a>).

Hill S, Pang T. Leading by example: a culture change at WHO. *Lancet* 2007;369:1842–1844.

Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Glasziou P, Guyatt GH. *Cochrane handbook for systematic reviews of interventions*. 11: Presenting results and 'Summary of findings' tables (available at <http://www.cochrane-handbook.org/>).

Golder SP, Loke YK, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Medical Research Methodology* 2006; 6:3.

Schünemann H, Brożek J, Oxman A, eds. *GRADE handbook for grading quality of evidence and strength of recommendation*. 3.2 [updated March 2009].

Assessing cost and resource implications

What is it about?

In recent decades there has been argument from a number of sources for the incorporation of economic considerations within guidelines. As health interventions are not free, people are not infinitely rich, and the budgets of health care programmes are limited, considering resources when developing recommendations highlights possible alternative use of interventions – defined as “opportunity cost”. It has been recommended that every set of clinical guidelines should include information on the cost implications of the alternative preventive, diagnostic and management strategies for each clinical situation. The stated rationale is that this information helps potential users to evaluate the potential consequences of different practices.

Several approaches to assessing the economic aspects of guidelines are used internationally by guideline development groups when considering the economic consequences of health interventions. One of the major challenges that applies to all of these approaches is how to deal with the high variability in costs within and across settings and over time.

Analysis of resource use and costs

An analysis of resource use (and, if appropriate, costs) is basic to any analysis of economic aspects associated with a guideline. Generally, all important resource use associated with the intervention – and, if available, the comparator – are assessed. This analysis has three steps, namely:

- identification (*what type of resource use is associated with this intervention?*)
- measurement (*how much of this is used?*)
- monetary valuation (*what does this cost?*)

The minimum requirement for WHO guidelines is to provide assessment of the identification of resource use.

Economic evaluation

A formal economic evaluation of cost-effectiveness takes into account the costs and health outcomes (effects) of an intervention assessed in relation to its comparator, and usually aggregates them in an incremental cost-effectiveness ratio (ICER). Effectiveness measures can be natural units (e.g. disease episodes or deaths prevented), two-dimensional quality-adjusted life years (QALYs) in a cost-utility analysis, or can be expressed in monetary terms in a cost-benefit analysis. Cost-effectiveness analyses often use decision-analytic methods in order to combine evidence from different sources and to extrapolate from the limited time-horizons of existing studies on health outcomes. Once the cost-effectiveness of an intervention is established, an evaluation should be made as to whether the intervention is affordable and represents value for money.

Assessing resource implications

Besides considerations of costs and the effects of alternative interventions, there are resource implications when a recommendation is implemented but the practice or policy changes in a specific setting. Important resource implications to consider include health system implications such as training and supervision requirements, referral support, equipment and infrastructure requirements, and monitoring and evaluation. In the GRADE approach, these considerations are generally part of the feasibility assessment when formulating recommendations.

What do I need to do?

Generally, it is advised to:

- include consideration of the economic consequences of your recommendations early, at the stage of defining the scope of the guideline;
- discuss and decide with the guideline group, ideally including a health economist, how to address explicitly the economic aspects of the guideline.

The following section describes the minimum requirements for any WHO guideline.

Qualitative description of resource use

For WHO guidelines, a summary of resources used by each intervention should be provided for the key recommendations and their alternatives (this is equal to "Identification of resource use" above as the first step of a costing analysis). The description of resource use and costs should be made from the perspective of the health system.

Researchers should define the main resources required for the specific intervention. The definition should include resource use associated with the provision of the intervention, subsequent investigations and care, and adverse effects. These should be grouped as costs incurred by the patient, the health system and society. Those incurred by the patient and health system should always be described (e.g. drug, admissions, visits, examinations). Other resources, such as patient and care-giver time, should generally be considered only when they are considered to be very important in that context as they are difficult to measure and to put a value on reliably. It is also important to define the time horizon for inclusion of resource use – when are important differences in resource use likely to occur (in the short term or the long term)? As a second step, if there is a comparator, differences in resource use between the intervention and the comparator should be included.

Assessing resource implications

Assessment of resource implications should include all relevant health system implications of implementation of the guideline – such as training and supervision requirements, referral support, equipment and infrastructure requirements, and monitoring and evaluation. At the minimum, a WHO guideline should include a qualitative description of the resources needed for implementation relative to current practice. A scenario approach could be used to account for different settings (Tan-Torres, 2006).

When the GRADE system is applied, this step should be undertaken at the stage of developing the recommendations (specifically when implementation and feasibility issues are considered when moving from evidence to recommendations).

The steps described in this section are the minimum for any WHO guideline and may be all that can be done if data are scarce. If appropriate, analyses can be taken further. In this case, the next section provides advice on possible steps to take.

What can be done?

The qualitative description of resource use described above can provide a more quantitative analysis if one or two more steps are included – i.e. measurement and monetary valuation of resource use. A full economic evaluation might be worthwhile if an unbiased effectiveness measure is available, and a review of existing economic studies may be useful to inform the definition of resource use (for a costing exercise only or for a full economic evaluation).

Analysis of resource use and costs – measurement

Once resource use associated with alternative interventions is identified, the next step is to measure resource use, and differences in resource use, between intervention and comparator. The evidence for these measurements may come from a variety of sources, and may differ in terms of quality. Sources of evidence include systematic reviews, randomized trials and observational studies, as well as local prescription or hospital databases. Where possible, resource differences should be derived from studies that directly compare alternatives; observational data (e.g. insurance claim databases) are less reliable because of biases. Problems with applicability of resource use data from trials may arise if the trial setting and time horizon do not adequately reflect resource use in practice.

Analysis of resource use and costs – monetary valuation

Once resource use is measured, a range of monetary values can be estimated for each item of resource use. Differences should be presented between cost of the recommended intervention and cost of the comparator (incremental cost). As WHO generally makes global recommendations, it is useful to present cost ranges for different health systems and, where possible, for different levels of income (e.g. low-resource versus high-resource settings).

For reporting on this costing exercise, it is important not just to document the aggregate costs (number of units of resource use x unit costs of resource) associated with an intervention, but also to report as far as possible disaggregated costing information (i.e. all the associated resource use and unit costs separately). Guideline users can then judge the applicability of costs to their setting and, if necessary, re-work the data with different monetary values from the ones used.

Review of existing studies

A review of existing economic studies may be useful to inform definition of resource use. However, the applicability of results from existing economic studies needs to be carefully assessed as these studies are very context-specific. For advice on how best to search for relevant economic studies, see chapter 15 of the Cochrane handbook (Higgins et al., 2008). Other tools that might be useful for the assessment of existing economic studies include the Checklist for Assessing Economic Evaluations (Drummond et al., 2005, chapter 3) and the Consensus Health Economic Criteria (CHEC) list (Evers et al., 2005).

Formal economic evaluation

Economic evaluations should be undertaken only if unbiased estimates of the effect of an intervention and its comparator, as well as cost data, are available. The technical requirements for these analyses are

generally substantial and commissioning of the work is recommended. If a modelled analysis is undertaken, all model assumptions and data sources for cost and outcome data should be made available (see Philips et al., 2004) for good practice criteria for models.

Example

The following example of a qualitative description of economic considerations, following a review of published studies, is from *Guidelines for the psychosocially assisted pharmacological treatment of opioid dependence* (WHO, 2009, pp 27–28).

In a recent study of opioid substitution therapy in different countries, resource-use and cost data relating to methadone and buprenorphine maintenance treatment were collected in selected WHO member states (Indonesia, Iran, Lithuania and Poland). The total monthly cost of providing long-term methadone and buprenorphine maintenance treatment (including an initial induction phase) ranged from as little as US\$ 26–36 in Indonesia and the Islamic Republic of Iran (approximately US\$ 1/day) to US\$ 296 in Poland (approximately US\$ 10/day). This provides an indicative range within which to locate the expected investment needed to provide methadone maintenance treatment to a service user in a low- or middle-income country. In high-income countries, costs for methadone and buprenorphine maintenance treatment are generally estimated to be US\$ 5000 per year, or US\$ 15 per day.

Estimation of the cost of providing medication is not an adequate basis for budgetary planning, because it may represent only a fraction of total service costs (e.g. <20% in the Islamic Republic of Iran). Studies in Australia, Canada, the United States and United Kingdom have estimated the impact of treatment on total health-care costs, social security costs, lost productivity and crime.

Thus, these studies estimate the economic “return on investment” in opioid-dependence treatment. They show that treatment of opioid dependence pays for itself, because savings in social costs are greater than the expenditure on treatment. It is difficult to extrapolate the results of these studies to lower-income countries. Estimates of cost-effectiveness in high-income countries have found that both methadone and buprenorphine maintenance are cost-effective, being well below accepted thresholds for cost-benefit analysis of treatment.

The cost of a “one off” episode of opioid withdrawal varies significantly between settings; it depends on the method of withdrawal, the length of treatment, the medication used and staff resources. Because of differences between maintenance and withdrawal, it is difficult to estimate the long-term cost implications of choosing between:

- *opioid agonist maintenance treatment, which is low intensity and long term, and has a low relapse rate;*
- *opioid withdrawal, which is high intensity and short term, and has a relatively high relapse rate.*

References and additional reading

Eccles M, Mason J. How to develop cost-conscious guidelines. *Health Technology Assessment* 2001;5(16):1–69 (available at <http://www.ncchta.org/fullmono/mon516.pdf>).

NICE. *The guidelines manual. Chapter 7. Incorporating health economics in guidelines and assessing resource impact*. London, National Institute for Health and Clinical Excellence, 2009, (available at <http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/GuidelinesManual2009.jsp>).

Drummond MF, Schulpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*, 3rd ed. New York, Oxford University Press, 2005.

Guyatt G, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A et al. Incorporating considerations of resources use into grading recommendations. *British Medical Journal* 2008;336:1170–1173.

Tan-Torres Edejer E. Improving the use of research evidence in guideline development. 11: Incorporating considerations of cost-effectiveness, affordability and resource implications. *Health Research Policy and Systems* 2006;4:2.

Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. 15: *Incorporating economics evidence*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008 (available at www.cochrane-handbook.org).

Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *International Journal of Technology Assessment in Health Care* 2005;21:240–245.

Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004;8(36).

Incorporating values and preferences

WHO needs to be explicit about values in a global context. [Values](#), the relative importance or worth of a state or consequences of a decision (outcomes relating to benefits, harms, burden and costs), play a role in every recommendation. Ethical considerations, concepts that determine what is right, also play a role.

The values used in making recommendations should reflect those of the people affected. Judgements should be explicit and should be informed by input from those affected (including citizens, patients, clinicians and policy-makers).

When differences in values may lead to different decisions or there is uncertainty about values, this should also be made explicit. If differences in values are likely to affect a decision, such that people in different settings would probably make different choices about interventions or actions based on differences in their values, global recommendations should specify which values were applied and should allow for adaptation after incorporating local values.

Additional reading

Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N et al. Systematic review of involving patients in the planning and development of health care. *British Medical Journal* 2002;325:1263.

Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J et al. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technology Assessment* 2004;8(15) (available at <http://www.nchta.org/fullmono/mon815.pdf>).

SIGN 50: a guideline developer's handbook. Section 4. *Involving patients and their representatives*. Revised edition 2008. Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

Formulating recommendations

A guideline panel is usually convened for 2–5 days to draft or review the guideline and the recommendations.

For each of the key recommendations, the quality of evidence should be made clear.

Recommendations should specify the perspective that is taken (e.g. individual patient, health care system or society) and which outcomes were considered (including which, if any, costs). The language used in recommendations should be clear and direct, indicating an unambiguous action (e.g. all patients with disease A should be offered treatment B by health professionals).

Where possible, the language should be consistent across recommendations (e.g. all strong recommendations phrased with “should”).

How a guideline development group decides on recommendations

Ideally the group should reach recommendations based on consensus. Consensus does not necessarily mean unanimity, however, and in some cases a vote may need to be taken. The group should discuss and agree on the process at the beginning of the meeting.

It is most effective if the group considers draft recommendations that have been prepared by the writing team. A suggested process is as follows:

- the draft recommendation is presented by the WHO staff, with a justification and reference to the relevant evidence summary;
- the evidence is reviewed and discussed by the group, considering the balance of evidence for benefits and harms;
- a first recommendation is agreed;
- the group considers costs, values and preferences;
- if necessary, the first recommendation is modified;
- final agreement on the recommendation is reached.

If the group cannot reach consensus, one option is to let the recommendation of the majority stand but to include a note in the guideline that records the minority view.

Additional reading

Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *International Journal for Quality in Health Care* 2003;15:31–45.

NICE. *The guidelines manual. Chapter 11. Creating guideline recommendations*. London, National Institute for Health and Clinical Excellence, 2007, pp 66–74 (available at <http://www.nice.org.uk/niceMedia/pdf/GuidelinesManualAllChapters.pdf>).

Dartnell J, Hemming M, Collier J, Ollenschlaeger G. Putting evidence into context: some advice for guideline writers. *Evidence-based Medicine* 2007;12:130–132.

SIGN 50: a guideline developer's handbook. Section 7. Forming guideline recommendations. Revised edition 2008. Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

Grading recommendations

The strength of a recommendation reflects the degree of confidence that the desirable effects of adherence to the recommendation outweigh the undesirable effects.

Desirable effects can include beneficial health outcomes, less burden and greater savings. Undesirable effects can include harms, greater burden, and increased costs. Burden here refers to the demands of adhering to a recommendation that patients or care-givers (e.g. family members) may find onerous – such as having to undergo more frequent tests or opting for a treatment that may require a longer time for recovery.

Although the degree of confidence is a continuum, the GRADE system defines two categories – strong and weak. A strong recommendation is one for which the guideline development group is confident that the desirable effects of adherence outweigh the undesirable effects. This can be either in favour of or against an intervention. A weak recommendation is one for which the panel concludes that the desirable effects of adherence probably outweigh the undesirable effects, but the group is not confident about the trade-off. Reasons for not being confident may include:

- absence of high-quality evidence;
- presence of imprecise estimates of benefit or harm;
- uncertainty or variation in how different individuals value the outcomes;
- small benefits;
- benefits that are not worth the costs (including the costs of implementing the recommendation).

Despite the lack of a precise threshold for moving from a strong to a weak recommendation, the presence of important concerns about one or more of the above factors make a weak recommendation more likely (see Table 3). Guideline development groups should consider all these factors and make the reasons for their judgements explicit.

Implications of a strong recommendation are:

- For patients: most people in your situation would want the recommended course of action and only a small proportion would not.
- For clinicians: most patients should receive the recommended course of action. Adherence to this recommendation is a reasonable measure of good-quality care.
- For policy-makers: the recommendation can be adapted as a policy in most situations. Quality initiatives could use this recommendation to measure variations in quality.

Implications of a weak recommendation are:

- For patients: the majority of people in your situation would want the recommended course of action, but many would not.
- For clinicians: be prepared to help patients to make a decision that is consistent with their own values.
- For policy-makers: there is a need for substantial debate and involvement of stakeholders.

Table 3. Factors that may influence the strength of recommendations

Factor	Examples of strong recommendations	Examples of weak recommendations
Quality of evidence	Many high-quality randomized trials have demonstrated the benefit of inhaled steroids in asthma	Only case series have examined the utility of pleurodesis in pneumothorax
Uncertainty about the balance between desirable and undesirable effects	Aspirin in myocardial infarction reduces mortality with minimal toxicity, inconvenience and cost	Warfarin in low-risk patients with atrial fibrillation results in small stroke reduction but increased risk of bleeding and substantial inconvenience
Uncertainty or variability in values and preferences	Young patients with lymphoma will invariably place a higher value on the life-prolonging effects of chemotherapy over treatment toxicity	Older patients with lymphoma may not place a higher value on the life-prolonging effects of chemotherapy over treatment toxicity
Uncertainty about whether the intervention represents a wise use of resources	The low cost of aspirin as prophylaxis against stroke in patients with transient ischaemic attacks	The high cost of clopidogrel and dipyridamole/aspirin as prophylaxis against stroke in patients with transient ischaemic attacks

Many recommendations are labeled as either strong or weak. However, because the “weak” label may sometimes be misinterpreted, other options exist. These include the use of terms such as “strong/conditional” or “strong/qualified”.

Additional reading

West S, King V, Carey TS et al. *Systems to rate the strength of scientific evidence*. Evidence Report/Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD, Agency for Healthcare Research and Quality, 2002. (available at <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.chapter.70996>).

Guyatt G, Gutterman D, Baumann M, Addrizzo-Harris D, Hylek E, Phillips B et al. Grading strength of recommendations and quality of evidence in clinical guidelines. *Chest* 2006;129:174–181.

WHO recommendations for the prevention of postpartum haemorrhage. Annex 3. Methods used for developing guidelines. Geneva, World Health Organization, 2007, pp 25–31 (available at http://whqlibdoc.who.int/hq/2007/WHO_MPS_07.06_eng.pdf).

Akl EA, Maroun N, Guyatt G, Oxman AD, Alonso-Coello P, Vist GE et al. Symbols were superior to numbers for presenting strength of recommendations to health care consumers: a randomized trial. *Journal of Clinical Epidemiology* 2007;60:1298–1305.

Research needs and priorities

A good guideline should summarize the available evidence as the basis of the recommendations. It is also good to highlight where the available evidence may be insufficient or inadequate. WHO guidelines should therefore identify research needs and, if appropriate, prioritize them.

In formulating research needs, guideline groups should be as specific as possible about what is needed and why. One format is EPICOT, described below (from Brown P et al., *British Medical Journal* 2006;333:804–806).

Suggested format for research on the effects of treatments

Core elements

- E Evidence (what is the current state of the evidence?)
- P Population (what is the population of interest?)
- I Intervention (what are the interventions of interest?)
- C Comparison (what are the comparisons of interest?)
- O Outcome (what are the outcomes of interest?)
- T Time stamp (date of recommendation).

Optional elements

- d Disease burden or relevance
- t Time aspect of core elements of EPICOT
- s Appropriate study type according to local need.

Peer review and plans for updating

The peer review process

WHO guidelines should undergo peer review during development and before the draft is finalized for publication. There are several stages at which peer review and external comment should be sought:

- Drafts of the questions formulated for the guideline should be circulated for comments to experts and end-users at WHO headquarters, regional offices and externally.
- If systematic reviews are undertaken or commissioned, the systematic review protocol (outlining search strategy and eligibility criteria) and included studies may be circulated to experts for comments on the methods and evidence identified.
- Drafts of evidence profiles and tables should be circulated to experts for identification of any missing evidence.
- Before the meeting, a draft of the document should be circulated widely for comment by experts and organizations representative of the relevant stakeholders.
- For full guidelines, a final draft with recommendations may be circulated for review before publication.

The process of reviewing comments and responding to them should be transparent. It is not necessary to respond to every single comment individually; this should be made clear at the beginning of the process. However, an “audit trail” should be drawn up to show how comments were handled, either as a version of the document with the changes, or as a separate summary.

If the guideline is circulated for comment after recommendations are finalized, be clear about what changes can be made. It is suggested that changes after finalization should be restricted to major errors of fact.

Different types of guidelines will have slightly different processes of peer review:

- Rapid advise guidelines: peer review can be limited to review of the complete draft only, immediately before final clearance, perhaps by 3–6 experts.
- Standard guidelines: more complete peer review would be expected, including:
 - review of questions;
 - review of evidence tables and completed draft recommendations (after the guideline meeting);
 - a record of the response to the comments and any changes that are made.
- Full guidelines: peer review would be expected to be as above, with additional review after a second draft.

Length of validity of guidelines

Guidelines will not last forever. WHO guidelines must be issued with a “use by” or “review by” date to provide an indication of how long the recommendations are expected to remain valid. There is no absolute

rule about the length of validity. In deciding on the date by which a guideline should be reviewed, take account of the pace of change of research on the topic, areas in which no evidence has been found, and the potential need for new advice. For standard and full guidelines, a minimum of two years and a maximum of five are suggested. There should also be a description of the department that will be responsible for initiating the review.

Additional reading

SIGN 50: *a guideline developer's handbook. Section 8. Consultation and peer review. Revised edition 2008.* Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

French SD, McDonald S, McKenzie JE, Green SE. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC Medical Research Methodology* 2005;5:33 (available at <http://www.biomedcentral.com/content/pdf/1471-2288-5-33.pdf>).

NICE. *The guidelines manual. Chapter 15. Updating guidelines and correcting errors.* London, National Institute for Health and Clinical Excellence, 2007, pp 86–89 (available at http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual_all_chapters.jsp).

Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T et al. *Updating systematic reviews.* Technical Review No. 16. AHRQ Publication No. 07-0087. Rockville, MD, Agency for Healthcare Research and Quality, 2007 (available at <http://www.ahrq.gov/downloads/pub/evidence/pdf/sysrev/sysrev.pdf>).

Shojania KG, Sampson M, Ansari TM, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine* 2007;147:224–233. <http://www.annals.org/cgi/reprint/147/4/224.pdf>.

Producing and publishing your guideline

Identify a writer early in the process. This can be a WHO staff member or an external writer contracted on a freelance basis. If the writing will be done by a staff member, it is important to estimate accurately the demands that will be made on the person's time. Once you have an idea of the approximate length of your document, you can make a rough calculation of the time needed and can begin negotiations with an external writer if necessary. WHO does not have a standard writing pay scale but WHO Press usually advises a minimum of US \$ 0.50 per word for writers, or a negotiated daily rate from current daily pay rates for consultants (available from the Human Resources Department). When negotiating fees and schedules, calculate a minimum of one week of full-time work to produce 5000 words.

It is strongly recommended to avoid the “committee” approach to writing a guideline. Asking experts to draft chapters for free may seem to be a cheap and efficient way of getting the job done, but unless you can guarantee quality, consistency and timely delivery, it will inevitably create more work than it eliminates.

You will also need an editor and a proofreader. WHO press maintains lists of approved freelance editors and proofreaders, and provides sample terms of reference and standard rates of pay for these tasks.

The best editors and proofreaders are often booked many months in advance, so plan production schedules as early as possible, and reserve their time accordingly.

Once you have a cleared, edited and proofed text, you will need to send it for layout. Again, WHO Press can advise on external typesetters and the specifications that you should include when contracting for this work. The WHO graphics team (GRA) also provides an internal layout service. As many design decisions have major implications for the cost of production, printing, dissemination and subsequent translations, it is worth discussing the possibility of using an existing publication template with WHO Press before engaging an external designer. You will need a cover design, an ISBN (international standard book number) and a barcode, the latter two issued by WHO Press.

Also give some thought to the forms in which your guideline will be disseminated, and in which format it will appear on the internet. At a minimum, you should contract your designer to produce a web-ready PDF – a smaller file size than the PDFs produced for print – that is easier to download and navigate. Depending on the length of the guideline and its intended audience, you may also wish to consider providing full-text HTML and additional materials, both electronic and paper. The WHO web team is a good source of advice on improving the impact of your content.

Internal print (PRT) will provide printing quotes and arrange for your files to be sent to the printer. You must have the printers' proofs checked again by your proofreader, so be sure to include this step in the initial proofreading contract. Once the print copies are delivered, you can focus on distribution and implementation.

Guideline format

All guidelines should have an executive summary, a main body and appendices. A general recommendation for the length of these sections, in particular for guidance to policy-makers, is the 1–3–25 rule – i.e. an executive summary of 1 page, the main guideline of three pages, and appendices of 25 pages.

The executive summary should contain the key recommendations of the guideline. As executive summaries are often read as stand-alone documents, the quality of evidence for each recommendation should be specified in the executive summary as well in the main body of the guideline.

Implementation and evaluation of impact

Implementation

Implementation of a guideline should be taken into account right from the beginning of its development. It is essential to decide on the desired outcome of implementation – is it to change clinical practice, to influence policy-makers, or does it have some other goal? A guideline project should ideally be nested within a departmental or other programme of work on the particular topic since that is more likely to lead to an effective plan for implementation. Implementation will generally be the responsibility of regions and national or subnational groups, which is why they need to be involved in the development of the guideline.

The basic steps for implementing a guideline are:

- Analyse local needs and priorities (look for additional data on actual practice).
- Identify all potential barriers and facilitating factors.
- Determine available resources.
- Design a strategy to support the adoption of the recommendations and to make the overall context favourable to the proposed changes (this will depend on where you are and what you are trying to do).

Additional reading

Grol R, Wensing M, Eccles M. *Improving patient care*. Edinburgh, Elsevier, 2005.

Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362:1225–1230 (available at <http://download.thelancet.com/pdfs/journals/lancet/PIIS0140673603145461.pdf?id=9d3ded37aa4dcc76:ffa1b50:121e8be73d1:-1a081245152834129>).

NHMRC. *How to put the evidence into practice: implementation and dissemination strategies*. Handbook series on preparing clinical practice guidelines. Canberra, National Health and Medical Research Council, 2000 (available at http://www.nhmrc.gov.au/publications/synopses/_files/cp71.pdf).

Wensing M, Wollersheim H, Grol R. Organizational interventions to implement improvements in patient care: a structured review of reviews. *Implementation Science* 2006;1:2 (available at <http://www.implementationscience.com/content/1/1/2>).

Implementation of CPGs to change practice and outcomes. In: Davis D, Goldman J, Palda VA. *Handbook on clinical practice guidelines*. Ottawa, Canadian Medical Association, 2007 pp 17–29 (available at <http://www.cma.ca/multimedia/CMA/Content/Images/ClinicalResources/PDF/English/CPGHandbook.pdf>).

SIGN 50: a guideline developer's handbook. Section 10. Implementation. Revised edition 2008. Edinburgh, Scottish Intercollegiate Guidelines Network, 2008 (available at <http://www.sign.ac.uk/pdf/sign50.pdf>).

Evaluation and monitoring

The guideline should include parameters or outcome measures that can be monitored for the main recommendations. Ideally, there should be baseline measures against which to assess performance in

relation to the targets of change of the recommendations. Follow-up measurements can be taken once the guideline is implemented. However, the level of detail for these indicators will depend on the target audience for the guideline and the plans for implementation.

WHO will develop further advice on guidance for implementation and evaluation.

Additional reading

Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Are audits wasting resources by measuring the wrong things? A survey of methods used to select audit review criteria. *Quality and Safety in Health Care* 2003;12:24–28 (available at <http://qshc.bmj.com/cgi/reprint/12/1/24>).

APPENDICES

These appendices contain useful resources frequently consulted by guideline development groups. Many other tools and references are available on the GRC web site (see <http://intranet.who.int/homes/rpc/grc/resources/>).

Appendix I. Effective guideline development group meetings

Some tips on running guideline development group meetings are listed below.

- Clearly lay out the scope of the meeting at the start, including:
 - what is expected from meeting participants;
 - what needs to be achieved during the meeting;
 - what can be done afterwards;
 - what follow-up will take place with meeting participants.
- Choose someone who is experienced in facilitating meeting processes to chair the meeting.
- Have a guideline development expert on the guideline panel.
- Keep to the agenda as much as possible; plan adequate breaks.
- Briefly review the previous day's work at the beginning of each day of the meeting.
- Management of conflict of interests:
 - collect declarations of interest well before the meeting so that there is enough time to intervene if necessary (e.g. if any invited participant needs to be excluded owing to major conflicts or to prevent there being too many participants with potential conflicts of interest);
 - at the meeting, have each participant verbally report potential conflicts of interest; present a draft statement of declared conflicts of interest that will be refined (with actions taken if necessary) and that will be presented again at the end of meeting for sign-off.
- Have someone write and project important decisions on screen (ie research questions, recommendations)
- If the purpose of the meeting is to sign off on questions for guideline development:
 - prepare a draft set of questions as formulated by the steering group;
 - circulate the questions ahead of time to all meeting participants.
- If the purpose of the meeting is to formulate recommendations:
 - distribute the evidence profiles before the meeting;
 - at the meeting, present draft recommendations that have been prepared by the guideline steering group (meeting participants will comment on these and refine them).

Appendix II. Critical appraisal of systematic review sheet

Please contact the GRC secretariat if you need explanation of how to use this appraisal sheet.

SYSTEMATIC REVIEW: Are the results of the review valid? ^e

What question (PICOT) did the systematic review address?	
What is best?	Where do I find the information?
The main question being addressed should be clearly stated. The exposure, such as a therapy or diagnostic test, and the outcome(s) of interest will often be expressed in terms of a simple relationship.	The <i>Title</i> , <i>Abstract</i> or <i>final paragraph of the Introduction</i> should clearly state the question. If you still cannot ascertain what the focused question is after reading these sections, search for another paper!
This paper: Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear <input type="checkbox"/>	
Comment:	
F - Is it unlikely that important, relevant studies were missed?	
What is best?	Where do I find the information?
The starting point for comprehensive search for all relevant studies is the major bibliographic databases (e.g., Medline, Cochrane, EMBASE, etc) but should also include a search of reference lists from relevant studies, and contact with experts, particularly to inquire about unpublished studies. The search should not be limited to English language only. The search strategy should include both MESH terms and text words.	The <i>Methods</i> section should describe the search strategy, including the terms used, in some detail. The <i>Results</i> section will outline the number of titles and abstracts reviewed, the number of full-text studies retrieved, and the number of studies excluded together with the reasons for exclusion. This information may be presented in a figure or flow chart.
This paper: Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear <input type="checkbox"/>	
Comment:	
A - Were the criteria used to select articles for inclusion appropriate?	
What is best?	Where do I find the information?
The inclusion or exclusion of studies in a systematic review should be clearly defined a priori. The eligibility criteria used should specify the patients, interventions or exposures and outcomes of interest. In many cases the type of study design will also be a key component of the eligibility criteria.	The <i>Methods</i> section should describe in detail the inclusion and exclusion criteria. Normally, this will include the study design.
This paper: Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear <input type="checkbox"/>	
Comment:	
A - Were the included studies sufficiently valid for the type of question asked?	
What is best?	Where do I find the information?
The article should describe how the quality of each study was assessed using predetermined quality criteria appropriate to the type of clinical question (e.g., randomization, blinding and completeness of follow-up)	The <i>Methods</i> section should describe the assessment of quality and the criteria used. The <i>Results</i> section should provide information on the quality of the individual studies.
This paper: Yes <input type="checkbox"/> No <input type="checkbox"/> Unclear <input type="checkbox"/>	
Comment:	
T - Were the results similar from study to study?	
What is best?	Where do I find the information?

^e Systematic Review Critical Appraisal Sheet, Oxford University 2005, Centre for Evidence Based Medicine. Available at: <http://www.cebm.net/index.aspx?o=1157> [accessed on June 15 2009]

Ideally, the results of the different studies should be similar or homogeneous. If heterogeneity exists the authors may estimate whether the differences are significant (chi-square test). Possible reasons for the heterogeneity should be explored.

The **Results** section should state whether the results are heterogeneous and discuss possible reasons. The forest plot should show the results of the chi-square test for heterogeneity and if discuss reasons for heterogeneity, if present.

This paper: Yes No Unclear

Comment:

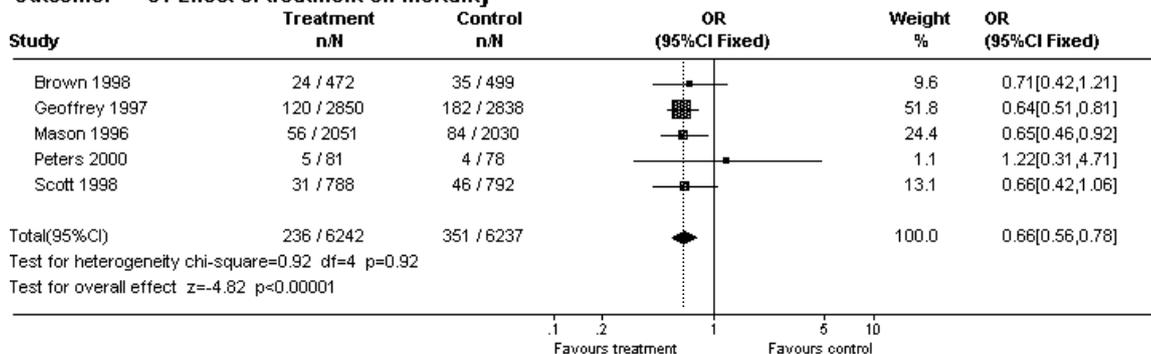
What were the results?

How are the results presented?

A systematic review provides a summary of the data from the results of a number of individual studies. If the results of the individual studies are similar, a statistical method (called meta-analysis) is used to combine the results from the individual studies and an overall summary estimate is calculated. The meta-analysis gives weighted values to each of the individual studies according to their size. The individual results of the studies need to be expressed in a standard way, such as relative risk, odds ratio or mean difference between the groups. Results are traditionally displayed in a figure, like the one below, called a **forest plot**.

Comparison: 03 Treatment versus Placebo

Outcome: 01 Effect of treatment on mortality



The forest plot depicted above represents a meta-analysis of 5 trials that assessed the effects of a hypothetical treatment on mortality. Individual studies are represented by a black square and a horizontal line, which corresponds to the point estimate and 95% confidence interval of the odds ratio. The size of the black square reflects the weight of the study in the meta-analysis. The solid vertical line corresponds to 'no effect' of treatment - an odds ratio of 1.0. When the confidence interval includes 1 it indicates that the result is not significant at conventional levels ($P > 0.05$).

The diamond at the bottom represents the combined or pooled odds ratio of all 5 trials with its 95% confidence interval. In this case, it shows that the treatment reduces mortality by 34% (OR 0.66 95% CI 0.56 to 0.78). Notice that the diamond does not overlap the 'no effect' line (the confidence interval doesn't include 1) so we can be assured that the pooled OR is statistically significant. The test for overall effect also indicates statistical significance ($p < 0.0001$).

Exploring heterogeneity

Heterogeneity can be assessed using the "eyeball" test or more formally with statistical tests, such as the Cochran Q test. With the "eyeball" test one looks for overlap of the confidence intervals of the trials with the summary estimate. In the example above note that the dotted line running vertically through the combined odds ratio crosses the horizontal lines of all the individual studies indicating that the studies are homogenous. Heterogeneity can also be assessed using the Cochran chi-square (Cochran Q). If Cochran Q is statistically significant there is definite heterogeneity. If Cochran Q is not statistically significant but the ratio of Cochran Q and the degrees of freedom (Q/df) is > 1 there is possible heterogeneity. If Cochran Q is not statistically significant and Q/df is < 1 then heterogeneity is very unlikely. In the example above Q/df is < 1 ($0.92/4 = 0.23$) and the p-value is not significant (0.92) indicating no heterogeneity.

Note: The level of significance for Cochran Q is often set at 0.1 due to the low power of the test to detect heterogeneity.

Appendix III. The PRISMA criteria for reporting of systematic reviews

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g. Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g. PICOS, length of follow-up) and report characteristics (e.g. years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g. databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e. screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g. piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g. PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g. risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. I^2) for each meta-analysis.	

Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g. publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g. study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g. sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g. healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g. risk of bias), and at review-level (e.g. incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g. supply of data); role of funders for the systematic review.	

Source: Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* 6(6): e1000097. doi:10.1371/journal.pmed1000097.