WHO Handbook for guideline development

March 2008, for update

This version of the Handbook will be updated in 2009. Your feedback and suggestions would be appreciated - please send them to grcinfo@who.int

Table of contents

Planning and scoping the guideline

Some definitions

A WHO guideline is any document containing recommendations about health interventions, whether they are clinical, public health or policy. The name of the document is not relevant.

Health information products that are NOT guidelines include:

- Documents containing standards for manufacturing health technologies such as pharmaceuticals, vaccines etc,
- 'how to' documents - for example, how to set up a particular type of research project, how to implement a particular type of service,
- documents that describe standard operating procedures for organizations or systems.

If you are not certain that your proposed document is a 'guideline' it should be submitted for review by the Guideline Review Committee.

If you are planning to produce a guideline, please consider the following types or products and decide which one best fits your purpose. The choice of product will determine the methods and time frame for development.

Emergency guidelines
An 'emergency guideline' is one that is produced in response to a public health emergency (for example emergence of a new SARS-type outbreak), where due to its mandate, WHO is required to provide global leadership and guidance. It is expected that this need to be produced quickly (indicative time frame 1-3 months), will be evidence-informed, but not supported by full reviews of the evidence and will be prepared mainly by the WHO staff responsible with limited external consultation or peer review. They should be published with a 'review-by ' date that indicates when they become invalid or when they will be updated or converted to a standard guideline.

Standard guidelines
A 'standard' guideline is one produced in response to request for guidance in relation to 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. It is not expected to cover the full scope of condition or public health problem. It will usually take between 9 and 12 months to complete, and should be prepared following consultation on the scope and question covered in the guideline, supported by systematic evidence reviews (that could be externally commissioned) and 1 or 2 consultation meetings. It may have a specified use by date depending on the expected rate of change of evidence in the topic area. It is expected that most WHO guidelines will fall into this group.

Full guidelines

A 'full' or management guideline is one that provides complete coverage of a topic or disease, such as guidelines on 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 be fully based on evidence reviews. It is likely to take between 2-3 years to complete, require several meetings of a guideline group, and should therefore be prepared only when WHO is the most appropriate agency to undertake this, or when there is likely to be no other group producing the guideline.

Books and other WHO publications

'Books' for this purpose are expected to be synthesis of recommendations from other sources, without necessarily containing new treatment recommendations. They are likely to be commissioned for production, not requiring evidence syntheses although being fully referenced, and requiring only limited consultation. Production times will vary.

Guidelines prepared in collaboration with other organizations

Guidelines for 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 a WHO document, and international society guidelines may have problems due to conflicts of interest in the funding for their development. The GRC will make case by case assessments of these type of proposals but, in principle, the following conditions will apply, (and noting the possible problems with copyright and ownership):

- adaptation or endorsement of another organisation's guideline should be initiated by the WHO Department concerned and not the external group
- that it could be considered when there was no existing WHO document or an outdated WHO document
- that minimum standards for WHO guidelines should be met, including no commercial entity funding, declaration and reporting of conflict of interest
- the approach for reviewing and summarizing evidence should be consistent with that recommended for WHO guidelines
- there should be global representation of experts in the development of the recommendations
- that the recommendations should be appropriate for a global audience.

Before you start

Before starting a guideline development process, you need to consider the following questions:

- **Why are you planning to do this?**
  - Who is asking for it? Is it a request from a member state? More than one? *WHO guidelines generally need to meet a global need, have a public health perspective and not duplicate existing resources.* If there is an existing product that meets the need, a new one is not required .
  - Why is this a priority? Again, WHO guidelines should meet global needs, and not duplicate existing resources. Is the topic one that *only* WHO can provide the best advice on? Is it required by the organization's governing bodies?
  - Is it part of a departmental program of work? Implementation of a guideline by HQ or by countries will be much easier if it fits with a program or project. If there is not program or project, 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 it to respond to poor quality practice or to try to change clinical practice or health policy? This should be the focus of most guidelines, and is what differentiates guidelines from textbooks or reference works. However, developing the guideline is only one step in the process/pipeline of implementing change in practice.
  - Is it in response to an emergency situation? *If so, please see the 'Emergency guideline' description.. These guidelines usually need to produced and published as fast as possible, ideally in 1-3 months and therefore the requirements and processes are different.*
  - Do you have agreement from your Director and ADG? You will need to get formal approval from both, so informal agreement at the outset is important.
  - Are there other departments who should be involved? Or who might be producing similar products? The answer to this is nearly always yes. ***Please*** - consult other relevant departments early in the process, decide which department has primary responsibility for the guideline and who will be involved in developing it.

If you can't answer all of these questions - we suggest you do not start!

Then consider and answer each of these questions:

- **Who is your target audience?**
  Most WHO guidelines need to 'speak' to 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 that meet the

need of policy makers, health care managers and clinicians simultaneously is not simple and should be avoided wherever possible.

- **When does it 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 some one to write it. If the guideline is going to cover a large number of questions, it will take up to 2-3 years to produce. Again, do you really need to do this??.

- **How much money is available?**
  For a standard WHO guideline, assuming that you will need to - contract a group or individual for preparation of evidence summaries (not systematic reviews), hold a single consultation meeting - pay for writing and editing and a small print run of the final document, please allow at least US$ 100 000. If you do not have this, or are not sure it will be made available - DO NOT START. Note that WHO may not accept money from commercial bodies for guideline development ad sources of funding for guidelines may need to be approved by Legal.

- **Are there existing guidelines documents that cover the same issue?**
  If so, what is the added value and justification for a WHO document? If the existing guidelines are from high quality national authorities (e.g. NICE, UK) and there is truly added value from a WHO version, 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 to a large extent by commercial entities, without consideration of conflicts of interest. If you want to start with these types of guidelines, you should assess the quality first. One option is to use the AGREE instrument: http://www.agreetrust.org/

- **What scientific evidence exists that can be used to guide recommendations?**
  Do you know of existing systematic reviews? If not, it is worth doing a preliminary literature search at this stage to try to get a sense of what information is available. For standard and full WHO guidelines you will need to have completed a systematic search for evidence prior to developing recommendations. If there is no evidence - what will be the basis of your guideline?

- **Who are key external organizations and experts, as well as stakeholders, who will need to be consulted or involved in the process?**
  It is worth spending some time generating a list at the beginning of the process, for several reasons. Firstly, you need to identify potential members of your guideline group. Secondly, there may be an additional group of experts or organizations who you may want to consult on the scope of the document, the questions it covers, and also the choice of important outcomes for decision making. Thirdly, there may be organizations and experts who could provide peer review of the completed draft guideline. These may include groups likely to oppose or criticize the output on the basis of scientific or philosophical differences, and while it may not be possible to reach agreement with them, it is

important that their input and comments should be considered. Fourthly, many of these groups and experts will be key to implementation of the recommendations in the guidelines and are more likely to help implement if they are involved from the beginning.

.

- **What guideline products do you want to produce?**
  WHO tends to produce printed documents as the default, which may not be the most useful output. Electronic versions may be more practical and cheaper, accompanied by short paper publications, wall charts, pamphlets etc.

- **Are you planning translations?**
  If so, which languages? Will you need technically qualified translators as well? (Budget implications here also!).

- **If you want to produce a 'book' please see:**
  http://intranet.who.int/homes/whp/write_edit/ for further advice.

For submission to the Guideline Review Committee you will need to produce a short outline and justification of your proposed guideline.

Scoping the guideline

'Scoping' the guideline is the process of defining the content, clinical questions and likely recommendations. If you get this stage right, the later stages are much easier!

1. Convene a small group of WHO staff to scope the guideline, including representatives of all involved departments (WHO coordinating/steering group).
2. List the priority topics for the guideline. If you are working on a clinical area (e.g. dengue) 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. 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 will need to be formulated, list these as well.
4. Do you really need to include all of these topics/questions/recommendations? The group should try to restrict the final list to the minimum at this stage, as it will inevitably expand during the guideline development!
5. Formulate the key question to be answered in the guideline, based on the topic list and possible recommendations. These questions will guide the evidence synthesis.
6. Once you have a complete scope that is agreed by the group, it should be circulated for comment to key external experts and organizations. (They need to be reminded that WHO is producing a guideline, not a textbook - as the responses will almost always try to expand the planned scope!)

7. Do a reality check after the feedback. Is what you are trying to do feasible? Is your time frame reasonable? Do you have the money to do it?

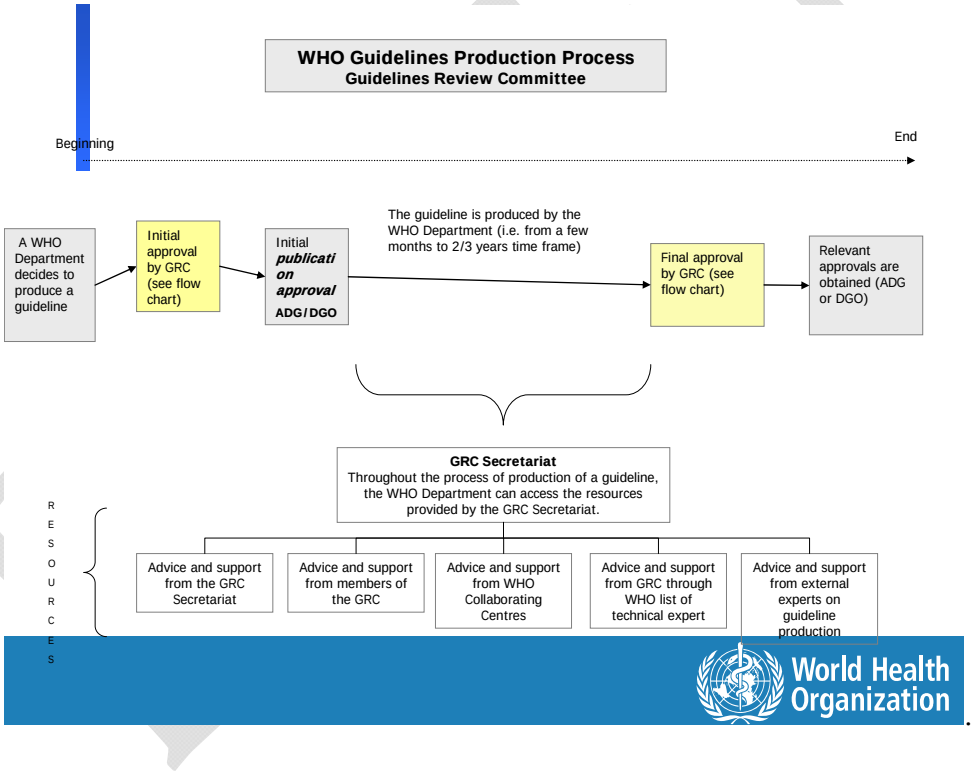
Additional reading

National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence. pages 12-15. Available from: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual__all_chapters.jsp

Section 1: Introduction. SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf

Approval

The approval process is shown in the figure below



The initial and final forms for publication clearance are at: http://intranet.who.int/homes/whp/publishing_process/approval/

Proposals need to be submitted to the GRC no later than a week before the date of the next meeting.

Guidelines groups - function and composition

It is usually necessary to convene a 'guideline group' to advise on the content of a WHO guideline. The guideline group should at a minimum develop and agree on the recommendations and review the completed document. A secondary purpose of a guideline 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 1 or 2 face to face meetings. The size, composition and function of the group are therefore key considerations in planning the budget for a WHO guideline.

Internationally, guideline organizations use their guideline development groups in different ways. Some involve them in all parts of the process including drafting text; others have them mainly involved in developing and endorsing the recommendations.

For WHO emergency guidelines, it would not be expected to convene a formal guideline group as the time frame for production of these guideline is too short. However, consultation either electronically or face to face with external experts would be ideal.

For standard and full WHO guidelines, a guideline group SHOULD be convened. It should include representatives from all regions likely to use the guideline including potential users, content experts and experts in the guideline development methods, including systematic reviews. Inclusion of end users is more likely to result in a guideline being appropriate to their needs and thus may contribute to successful implementation.

The group should be multidisciplinary and include:

- Content experts for all specialties involved
- Methodologists (experts in assessing evidence and/or developing guidelines; health economist, statisticians if appropriate)
- Representatives of potential stakeholders such as managers and other health professionals involved in the health care process who are likely to be 'end users' of the guideline
- Patients/consumers, although this can be difficult at a global level (see below)

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 in the guideline should not be chair of a guideline group. In most situations groups work most effectively if the Chair has some knowledge of the content, but is particularly an expert in facilitating groups and interpreting evidence. One option is to have a WHO staff member co-chair the group, and to ensure that the Chair does not have a 'veto' within the group.

'Consumers' involvement in WHO guideline development, although challenging at a global level, should be considered and encouraged from the start of the guideline process. It is particularly relevant for guidelines that have a predominantly clinical focus.

There are an increasing number of consumer groups operating at the international level. Many countries have non government organisations with members who might be available. For instance, in the patient safety area, WHO works with Patients for Patient Safety, a network of globally linked patient advocates and groups see: http://www.who.int/patientsafety/patients_for_patient/en/. There are a number of good consumer organisations in the medicines area. Benefits of involving consumers in guideline groups include formulating questions relevant to consumers, ensuring that relevant aspects of the experience of ill health are covered in planning effective treatment, for example, issues of communicating evidence-based choices to patients, providing information on the condition for self-care. Other important contributions include identifying and prioritising outcomes of importance, understanding the constraints on people in terms of being healthy and for input to preparing a consumer-accessible version of the guidelines. Barriers to consumer participation include lack of suitable consumer group, or time constraints, complexity of scientific language and language used by committee. In some cases, such as an illness-related group, illness may preclude participation - this has been one of the factors that has hindered consumer participation.

Experience from organisations such as NICE suggests that if consumers are going to be active contributors to guidelines groups, there needs to be more than one in the group and some training before hand is helpful . See NICE website for further information: http://www.nice.org.uk/

Gender representation and balance should also be considered in selecting group members.

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

- To advise on the choice of important outcomes for decision-making and developing recommendations
- 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.

In some circumstances, group members may contribute to writing sections of the guideline - this needs to be managed in such as way as to ensure a standard process and approach.

The size of a guideline group should be such to allow effective group interaction, but if possible, adequate representation and membership. Fifteen to 20 is feasible (and affordable!) Larger groups may become very difficult to manage. For additional consultation, electronic communication may be appropriate.

Representatives of commercial organizations may not be members of guideline groups. They may be invited to attend part of the meeting as observers, but should not attend the meeting when recommendations are being formulated.

Additional reading

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

Forming and running a guideline development group. 9. Making group decisions and reaching consensus. In: National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence. 2007. pag 20-9, 56-9. Available from: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual__all_chapters.jsp

Section 5: The guideline development group. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf

Conflict of interests

There are two issues to consider in relation to conflict of interest: declaration and reporting, and managing reported conflicts of interest in the guideline group meeting.

Declaration and reporting

Anyone invited to participate **in any way** in the development of a guideline must complete a Declaration of Interests, and must agree to the publication of their declaration in the guideline. The current WHO form for Declaration of Interests is on the web at:
http://intranet.who.int/homes/whp/pprg/pubpolicies/declaration_of_interests.shtml

No one may contribute to the development of a guideline until their Declaration of Interest has been reviewed by WHO, including if necessary the Legal Department. In the case of a member of a guideline group the completed form must be received by WHO at least six weeks before the first meeting of the group. If a proposed participant does not declare any conflict of interest they may begin to contribute. If any potential conflict of interest is declared the Declaration will be reviewed by LEGAL.

LEGAL 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 only to some areas of the guideline development group's work. In this case the participant will be absent when a committee considers these areas, and will not have access to the relevant documents.
- the conflict of interest is such as to preclude participation.

In all cases LEGAL's advice will be determinative.

Declarations of Interest for all participants in the development of a guideline should be provided to all other participants.

Anything that could be perceived to be a conflict of interest by any user of the guideline **must be reported**. Declared conflicts of interest should be reported following 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 which produce antibiotics.*
>
> *Dr Acar: he 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 with, and conducts occasional consultancies for, two pharmaceutical companies which produce antibiotics, and conducts occasional consultancies 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 adopted"*

The text in which conflicts of interest are reported **must be approved** by LEGAL prior to final review by the GRC.

Additional reading

Choudhry NK; Stelfox HT; Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002;287:612-17

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

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70

Managing the conflicts of interest during the guideline group meeting(s)

At each meeting of the guideline group, all members should provide a verbal summary of their written declaration of interest. 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 about this should be documented.

What constitutes a potentially significant conflict of interest is a matter of judgment. Some examples that could be considered significant are:

- owning in shares in a company that manufactures a product/technology that may be recommended for use in the guideline,
- having a family members that works for a company that manufactures a product/technology that may be recommended for use in the guideline,
- having a major academic program of work to do with a product or technology likely to be considered in a recommendation,
- receiving grants form a company with an interest in a specific product.

There are no absolute rules about how best to manage conflicts of interest in decision-making. Options include excluding the person from participating in discussion of a particular item, excluding the person from the part of the meeting that formulates the recommendation or taking no action at all. However, 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 any products or related products to those being considered at the meeting, and also holding shares in Biota. He, therefore, excluded himself from discussion of the late item on antivirals.*
>
> *Dr M.R. reported having been a consultant for Roche about drug research and development and is currently a member of a data safety and monitoring board for them; receiving royalties through the NIH on the use of gossypol for cancer, being a consultant to several start upcompanies 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 a family member being 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.*

All actions taken should be documented in the guideline.

Additional reading

Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, et al. Consensus development methods and their use in clinical guideline development. *Health Technol Assessment* 1998; 2(3). Available from: http://www.ncchta.org/fullmono/mon203.pdf

Formulating questions and choosing outcomes

Guidelines are a way of answering questions about clinical, communication, organisational or policy interventions, in the hope of improving health care or health policy. When you are planning your guideline, it is therefore helpful to structure it in terms of the questions that you will be answering.

One way of approaching this is to divide 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 3 categories.

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

- specify the purpose of the guideline - general and specific objectives may be helpful,
- draft the foreground and background questions as researchable questions
- prioritize these questions
- determine which questions will need evidence reviews.

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

The types of questions covered to start with will probably be a long list:

- What are the phenomena associated with the problem?
- What is frequency of the problem?
- What causes the problem? (aetiology)
- Does this person have the problem? (diagnosis)

- What happens if you get the problem? (prognosis)
- How can we treat the problem?

To turn these into answerable questions, we suggest you use the 'PICOT framework:

Population *In women without HPV*	(What factors are essential?)
Indicator/**I**ntervention *Does HPV vaccine*	(Specific intervention or class?)
Comparator *Compared with no treatment*	(Compared to nothing or standard treatment)
Outcome *Reduce rates of CIN*	(Patient relevant outcomes?)
Time *For life?*	(Short-term or long term?)

This format cans be used - with slight modifications - for prevalence and incidence questions and diagnostic questions as well:

- *In women in Uganda (P), what is the frequency of breast cancer(O) each year (T)?*
- *In babies born to HIV positive women (P) does screening with a new rapid diagnostic test (I, C) prior to the age of 6 week (T) accurately detect disease?*

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

- **What are the phenomena/experiences/views that influence patient behaviour?**
  - *Observational studies (e.g., qualitative research)*
- **What is frequency of the problem? (FREQUENCY)**
  - *Random (or consecutive) sample*
- **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)**
  - *Randomised controlled trial(s)or systematic reviews of trials?*

Other possible guideline questions are:

- **How is the effective treatment best delivered (organizational intervention)?**
  - RCTs or systematic reviews of RCTs

- **How are the pros and cons of treatment best communicated with the patient or family member (communication intervention or strategy)?**
    - RCTs or systematic review of RCTs)

The draft questions should be sent to an international panel of clinicians, researchers, programme managers and consumers for comment and revision. Panel members should be asked to identify omissions from the key questions to be covered by the guidelines, or relevant outcomes to be considered in developing recommendations.

Additional reading

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

Akobeng AK. Principles of evidence based medicine. Arch Dis Child 2005;90:837–40.

Choosing and rating outcomes

In addition to reviewing the questions, the panel should be asked to identify the key outcomes that need to be considered in making recommendations. The purpose of this is to: (1) identify the outcomes that will be critical for making decision and recommendations, (2) identify the data that should be sought from the process of evidence retrieval and synthesis.

The WHO steering group should make an initial list of possibly relevant outcomes, including both desirable and undesirable effects. Members of the panel should be asked to score the relative importance of each outcome from 1–9, where 7–9 indicates the outcome is critical for a decision, 4–6 indicates it is important and 1–3 indicates it is not important. The average of scores for each outcome can be used for determining the relative importance of each outcome although it is helpful to have the range of results provided as well. The panel should also be asked to identify additional important outcomes not included in the list of potential outcomes identified by the WHO steering group.

If necessary, the final rating of outcomes can be reviewed and confirmed at the guideline group meeting.

Additional reading

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 Infect Dis* 2007;7:21-31

WHO Recommendations for the prevention of postpartum haemorrhage. Geneva: World Health Organization. 2007. pages 6-8. Available from: 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 an essential component for developing or to inform a recommendation. The summary of research evidence should be based on all published studies, and where relevant, unpublished ones. This is a systematic review. Quantitative synthesis of the results of studies found through a systematic review to estimate the overall effect of an intervention is a meta-analysis. This step may or may not be present or feasible or useful. In comparison with traditional narrative reviews, systematic reviews reduce the risk of selective citation and can be of help in detecting publication bias.

For guideline development, existing reviews should be used wherever possible and updated if needed. However, in some instances a published review may turn out not to be relevant to the question for the guideline. If there are key questions to be covered in the guideline that are not answered by an existing systematic review, consideration should be given to undertaking or commissioning a new review. Preparing systematic reviews can take over a year and requires capacity and resources. *We suggest 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.*

Prioritizing evidence retrieval

Trying to retrieve evidence to support every recommendation in a guideline may simply not be feasible. This is where the process of identifying priority questions or issues for the guideline to address becomes important. (see section on questions)

For WHO guidelines, it is suggested that controversial questions, or those where there is need to change practice, should be supported by evidence summaries. It is less critical to provide a full review , for example, for a section on the epidemiology or pathology of a disease, if this is provided as background information only.

Evidence retrieval

Before starting a major process of evidence retrieval for a WHO guideline, we suggest that you start with a search for existing guidelines. If they are of reasonable quality, they should have lists of references, including systematic reviews, that can be used as a 'short-cut into the search.

The second step in evidence retrieval is identifying relevant systematic reviews for each question. This can be done in-house, or can be contracted to a group preparing the evidence summaries. Systematic reviews can include all types of study design, and it is worth considering what is most relevant to your questions first. If you choose to do the search and retrieval in-house:

- There needs to be a clearly documented search strategy for guidelines/systematic reviews/primary research that specifies 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 applied, with the date the search was conducted and/or updated, should be documented in the final guideline.

The most readily accessible biomedical database is PubMed. By using the PubMed 'Clinical Queries' or 'Special Queries' options, you can set up specific searches to identify systematic reviews or different types of studies. This includes searches of the Cochrane Database of Systematic Reviews. Systematic reviews of policy interventions (such as pricing of pharmaceuticals) can be difficult to find and other search strategies will be needed. The Campbell Collaboration lists studies of effectiveness of social and educational policies and practices. *When in doubt - get help.*

Once you retrieve the reviews, you need to check two things - date of last update and quality. For existing reviews, standard criteria, such as the checklist developed by the Oxford Centre for Evidence Based Medicine see link: http://www.cebm.net/index.aspx?o=1157, should be used to critically appraise them, together with an assessment of the relevance of the review to the questions being asked.

Generally, if the review is more than 2 years since its last update, there needs to be a search of primary literature to identify new relevant studies. This is more complex and time consuming and you should seek advice from an information specialist. Depending on the question to be addressed the following types of studies may be useful:

- **What are the phenomena/thoughts?**
  - *Observational studies (e.g., qualitative research)*
- **What is frequency of the problem? (FREQUENCY)**
  - *Random (or consecutive) sample*
- **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)**
  - *Randomised controlled trial(s)*

For a WHO guideline, it is important to try to include studies from the developing world as well as the more standard literature. Developing country journals are not well represented in PubMed and commercial databases such as EMBASE, CAB Abstracts. Regional databases (AIM, IMEMR, HELLIS; LILACS, WPRIM) grouped under the general heading of the Global Health Index http://www.who.int/ghl/medicus/en/ contain unique citations and in many cases full text articles. Regional offices of WHO have supported the development of the these indexes to highlight the health research literature of developed world. The majority of journals indexed by Regional databases are not indexed PubMed and other databases. The WHO Library can also suggest other databases to search depending of the topic areas (Legal, Social Sciences, Political etc)

Retrieval of 'grey literature' such as Ministry of Health reports, case studies and unpublished studies, is best done based on the results from journal article searching from PubMed, Regional databases and other sources. Terms such as the key authors and institutions can be used with Google or other search engines to identify grey literature cited on the internet. It is also important to scan key web sites individually as general search engines are not capable of retrieving all the relevant information on a web site. Obtaining training in using search engines is important to limit research results to pertinent information. The combination of efficient use of search engines and targeted website/authors is much more effective for identifying unique information than large unfocused searches. Personal contact with key informants will help to identify sources of information not found in the published journals or cited on website.

Additional reading

Finding the Evidence - Literature Searching for Guideline Development. In: New Zealand Guidelines Group (NZGG). Handbook for the preparation of explicit evidence-based clinical practice guidelines. Wellington: New Zealand Guidelines Group (NZGG). 2001. pag 23-37. Available from: 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 Info Libr J* 2008;25:23-30

Evidence assessment

This step is the main preparatory work for the work of the guideline group and the formulation of the recommendations. The basis of the evaluation of evidence is the systematic review(s) of available studies identified from the process of evidence retrieval. To enable a guideline group to 'digest' the evidence, summaries need to be prepared and it is easiest to do this using a tabular format.

For interventions, evidence profiles should be created for the priority questions using the GRADE approach, to allow a structured and transparent judgment of the quality of evidence for each outcome rated as important by the guideline panel. These tables are based on systematic review(s). For each question, data should extracted for all of the outcomes (benefits and harms) that were rated to be important . If there is more than one systematic review, start with the best one, and supplement it as needed with additional data from other good quality systematic reviews. If the published systematic reviews are not recently updated (within the last 2 years), a search should be conducted to retrieve more recent studies or studies in press that could be relevant to the specific issue.

Diagnostic interventions will shortly be included in the GRADE system.

For summaries of studies of the epidemiology of a condition, or prognostic studies, or observational data use a study by study table. Draft evidence tables should be sent to the members of the guideline group prior to the meeting with a request for identification of any missing relevant evidence.

Additional reading

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 (NHMRC). 2000 Available from: http://www.nhmrc.gov.au/publications/synopses/_files/cp69.pdf

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MB, Contopoulos Ioannidis DG, Lau J. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001; 286:821-30.

Assessing and Applying the Evidence. In: New Zealand Guidelines Group (NZGG). Handbook for the preparation of explicit evidence-based clinical practice guidelines. Wellington: New Zealand Guidelines Group (NZGG). 2001. pag 42-54. Available from: http://www.nzgg.org.nz/download/files/nzgg_guideline_handbook.pdf

Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, Wells S, Glasziou P, Heneghan C. The GATE frame: critical appraisal with pictures. *Evid Based Med* 2006 ;11:35-8

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. *Ann Intern Med* 2007;147:871-5.

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 from: http://www.has-sante.fr/portail/upload/docs/application/pdf/eval_interventions_ameliorer_pratiques_guide.pdf

Cochrane Handbook for Systematic Reviews of Interventions. 8 Assessing risk of bias in included studies. Higgins JPT, Altman DG eds. Preliminary draft, 3 December 2007. Available from: http://www.cochrane.org/resources/handbook/Handbook5_Bias_V10.pdf

Using GRADE

The GRADE approach has 2 main steps:

1. Evaluation of the quality of evidence.

Quality is defined as the 'extent to which one can be confident that an estimate of effect or association is correct' . It is a continuum; any discrete categorization involves some degree of arbitrariness. It is based on the following criteria:

- study design and any limitations of the 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 the populations, interventions and settings where the proposed intervention may be used

- the likelihood of publication bias

And additionally for observational studies:

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

'Quality' of evidence is categorized as 'high', 'moderate' , 'low' or 'very low' and the definitions are shown below.

Quality of evidence and their definitions

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

2. Preparation of a summary of findings - the results of the studies, using both relative and absolute measures.

Preparing GRADE tables

Start with the GRADE template table. This can be done in WORD or using the GRADE PROFILER software: http://www.gradeworkinggroup.org/members/balance.htm For support on this please contact hjs@buffalo.edu.

GRADE tables are constructed by 'rows' for each outcome. You will need at least one table per question and to make the Table more informative and readable, beneficial outcomes should be separated from harms/side-effects.

In the first row, fill in the most important beneficial outcome (in the PPH example, this is mortality).

Identify the systematic review(s) that include studies reporting the relevant outcomes. Not all studies in the reviews may report the outcome you are interested in -for each outcome, data should be presented from the subset of studies in the review that reported it.

Fill in the column, 'number of studies' - this is the number of studies in the review that report the outcome. For future reference and checking, we suggest you list these studies as a footnote to the table.

You then need to:

1. assess the quality of the evidence for these studies,
2. extract summary results for the second half of the GRADE table - relative and absolute measures of effect or where continuous outcomes are reported, the summary estimate of effect (weighted mean difference or standardized mean difference, and variance).

Quality of evidence evaluation

To complete a GRADE table for quality assessment, you need to consider the following:

- 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).

These are the headings of the columns of the GRADE table.

1. Study design

Studies are broadly classified as 2 types:

1. RCT – randomized controlled studies or randomized cluster trials
2. 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 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 RCTs begins as 'high' quality evidence and evidence from observational studies begins as 'low' quality evidence.

2. Limitations

Having defined the type of study(ies) you need to consider how well they were performed and analysed. For randomized controlled trials (RCTs), 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

There are many checklists for assessing quality of RCTs. This is a minimum set of criteria. If you want to find out more about assessing quality of RCTs see "How to evaluate the quality of RCT" for additional references.

For observational studies, the main criteria depend on the design: case control or cohort studies. For both, 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 needs to be evaluated. For a checklist for evaluating observational studies, see the Newcastle-Ottawa checklist and its Manual or GATE.

Rating the limitations in study design requires a degree of judgment. You need to decide whether the studies have 'no limitations' or 'serious limitations' or 'very serious limitations'.

'No limitations' generally means that the majority of studies meet all of the minimum quality criteria for the design. The implication of this is that the rating of quality of evidence remains the same as the starting assessment.

'Minor limitations'
Sometimes minor flaws are found when analysing how the available studies were designed and performed. For example, allocation concealment may not be reported for 1 study out of several in a systematic review or a study could be non blinded , but report objective outcomes. If you decide there are minor limitations, these should be noted in a footnote, but do not usually downgrade the quality.

'Serious limitations'
Serious means that one of the minimum criteria for quality is not met by the majority in studies in the review. This results in a '-1' score for the overall quality rating (e.g high become moderate).

'Very serious limitations'
Very serious means that at least 2 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 actually used for downgrading the quality of evidence and the reason for the assessment should be explained in a footnote for the table.

If most of the evidence for an outcome (based on the weight given to each study in the meta-analysis) came from trials that did not have serious limitations, the overall

assessment for that outcome will be that there are no important limitations and so the final judgment on the quality of evidence will coincide with the study design.

3. Assessing consistency

Consistency refers to the similarity of estimates of effect across studies. To evaluate the degree of consistency of the results of the available studies you should evaluate the *direction* and *size* of the effect for each outcome. If a formal metanalysis was conducted the result of the test for heterogeneity can be used to help assess consistency. Variability or inconsistency in results can arise from differences in the populations in the studies, differences in the interventions, or outcomes.

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 important inconsistency. If there is inconsistency in the results, such as the largest trial showing results that contradict smaller trials, then a '-1' score should be rated. If the results are very heterogeneous, choose 'very serious' and this will downgrade the evidence for this outcome by 2 levels.

In case only one study is present, consistency is not applicable as a criteria.

4. Assessing directness

Directness or generalisability or external validity of study results or applicability are all synonymous. There are two types of indirectness.

1. Indirect comparison – occurs when a comparisons of intervention A versus B is not available, but A was compared with C and B was 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.
2. Indirect population, intervention, comparator, or outcome – the question being addressed by the guideline panel or by the authors of a systematic review is different from the available evidence regarding the population, intervention, comparator, or an outcome.

To determine whether important uncertainty exists, you should 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 judgments on directness should not be excessively stringent.

For some therapies—for example, 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 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 dimensions and has practical relevance for the implementation of study results in actual practice. Examples of uncertain directness in WHO recommendations are (avian flu, opioid dependence). The judgment 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, for example, in a developed world setting, and the intervention is likely to be altered according to setting, this would be rated as 'major uncertainty' (and therefore scored as -2).

5. 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 due to resulting uncertainty in the results.

For dichotomous outcomes, downgrade the quality of evidence for any of the following reasons:

1. total (cumulative) sample size is lower than the calculated optimal information size (OIS)

2. total number of events is less than 300

3. 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect includes no effect and
    a. if recommending in favor of an intervention – the upper confidence limit includes an effect that, if it were real, would represent a benefit that, given the downsides, would still be worth it
    b. if recommending against an intervention – the lower confidence limit includes an effect that, if it were real, would represent a harm that, given the benefits, would still be unacceptable

4. 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect excludes no effect but
    a. if recommending in favor of an intervention – the lower confidence limit crosses a threshold below which, given the downsides of the intervention, one would not recommend the intervention
    b. if recommending against an intervention – the upper confidence limit crosses a threshold above which, given the benefits of an intervention, one would recommend the intervention.

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. Under such circumstances one may not downgrade the quality of evidence for imprecision.

For continuous outcomes, downgrade the quality of evidence when:

1. 95% confidence interval includes no effect and the upper or lower confidence limit crosses the minimal important difference (MID), either for benefit of harm
2. if the MID is not known or use of different outcomes measures required calculation of an effect size (ES), downgrade if the upper or lower confidence limit crosses an effect size of 0.5 in either direction.

6. Other considerations

Publication/reporting bias

Publication bias is a systematic underestimate or an 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 prototypical situation that should elicit suspicion of reporting bias occurs when published evidence is limited to a small number of trials, all of which were funded by a for-profit organization.

You should consider the extent to which they are uncertain about the magnitude of the effect due to selective publication of studies or reporting of outcomes and 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 be confident about the results. In those situations, the weak study design is unlikely to explain all of 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 becomes the evidence.

Magnitude of effect	Effect measure	Quality of evidence
Large	RR >2 or <0.5	upgrade 1 level
very large	RR >5 or <0.2	upgrade 2 levels

(RR - relative risk)

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 but this only applies to studies not downgraded for any reason. To rate the presence of dose-response gradient:

- If there is no evidence of dose-response gradient , there is no change
- If there is evidence of 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 be working to underestimate the true treatment effect. For instance, if only sicker patients receive an experimental intervention or exposure, yet they still fare better, 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»

The summary of rating criteria is in the table below. The final score is determined by adding the additional ratings to the original baseline score.

Table. GRADE quality of evidence: assessment criteria

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

2. Summary of findings

The second half of a GRADE table is the summary of findings. For each outcome, you will need to extract the results from the review(s).

The following information is needed for dichotomous outcomes:

- total number of patients in each group
- 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%CIs).

We suggest that wherever possible, you use the results as presented in the review rather than recalculating, assuming that the review methods are correct. That said, you may have to calculate absolute effect sizes and NNTs if you wish to represent these to the guideline group.

For continuous outcomes you will need:

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

It is advisable that one reviewer extracts data from the systematic reviews and/or from single studies and prepares drafts of the GRADE evidence profiles with detailed footnotes explaining the judgments that were made. Each judgment should be made

explicit and available to the reader in order to increase the transparency of the whole process. These should be checked by at least one other member of the team.

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. *Am J Respir Crit Care Med* 2006;174:605-14

Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schunemann H. An emerging consensus on grading recommendations? Evid Based Med 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 from: http://www.who.int/making_pregnancy_safer/publications/PPHTables2.pdf

Hill S, Pang T. Leading by example: a culture change at WHO. *Lancet* 2007;369:1842-4

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. Preliminary draft, 16 October 2007. Available from: http://www.cochrane.org/resources/handbook/Handbook5_Presenting_V6.pdf

Assessing evidence for other types of questions

1. The use of GRADE for assessing evidence for diagnostic tests is being developed.
2. Assessing quality of observational studies for question of prognosis and risk factors can be done using the criteria for observational studies described in section 'Study Limitations'.
3. Evidence such as countries case studies should only be described, and not assessed for quality. ¨

Additional reading

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

Assessing resources use and costs for guidelines

Why incorporate economic considerations in WHO guidelines?

Over last 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 in the development of recommendations highlights

possible alternative use of interventions, defined as "opportunity cost". It is a universal phenomenon, for all type of health systems. In addition, 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.

At present there are several methods for incorporating economic considerations that are used by guideline development groups when considering economic consequences of treatments, but there is no consensus yet on the best approach. One approach is to determine whether an intervention is 'cost-effective' and then whether it is 'affordable'. However, the practicalities of making these judgements are not straightforward particularly for a global audience. Some methods for producing 'cost-conscious' guidelines are outlined below.

How can economic aspects be explicitly taken into account in WHO guidelines?

There are three main approaches that are possible:

- a basic description of resources and costs for each intervention considered and its comparator if appropriate; and/or
- a review of existing economic studies for each recommendation; and/or
- a formal economic evaluation (including economic modelling)

These are described below. The minimum for any WHO guideline is to provide a description of resources and costs for key interventions recommended in the guideline. For further help and advice on this aspect, contact Dr David Evans or Dr Tessa Tan-Torres.

Describing resource use and costs

A summary of resources used for each intervention should be provided for the key recommendations in each guideline. If there is a comparator, differences in resource use should be included. The presentation of resources differences should follow the standard method for the definition of costs: identification, measurement and monetary valuation. For WHO guidelines, this should usually be based on the perspective of the health system.

Firstly, researchers should define the main resources and costs required for the specific intervention. These should be grouped as those incurred by the patient, the health system and society. Those incurred by the patient and health system should be always considered and described (e.g. drug, admissions, visits, examinations). Other resources and costs, such as patient and caregiver time should generally be considered only when they are considered very important in that context as they are difficult to measure and value reliably.

Secondly, the difference in resources for each alternative considered in the question should be presented. Where possible, resource differences should derive from studies directly comparing alternatives: observational data (e.g. insurance claim databases) are less reliable because of the biases in these data.

Thirdly, an estimate of the range of monetary values should be provided. 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. The incremental costs are preferable to average costs for decision-making. Explicit presentation of the resources and costs potentially permits simple reworking of data with different monetary values from the ones used.

Review of existing economic studies

Identifying and assessing economic studies may or may not be informative for a guideline. The limitation of economic studies generally is that they are usually very context specific, and can be applied only in the setting from which they are derived. Studies that are analyses of costs can be used to inform the definition of resources used in resources table summaries. Published cost effectiveness analyses need to be evaluated very carefully before being accepted as valid; many studies show that published economic evaluations are likely to be flawed or biased. In addition, they are generally limited in applicability so unless full access to the model is available, they are unlikely to be useful.

Formal economic evaluation

Formal economic evaluation (that is cost-effectiveness evaluation and its variants) has been recommended as an important tool for decision making. However, there are several problems with incorporating economic evaluations into the development of recommendations at a global level. Firstly, the quality of economic evaluations depends on the availability of unbiased estimates of the effect of an intervention and its comparator as well as costs. These can be difficult to obtain and may require substantial primary research. As economic evaluations are designed to provide an estimate of the incremental differences between alternatives, effectiveness studies and their careful quality evaluation are key.

Secondly, interpreting the results of an economic evaluation requires a context for assessing whether an estimate of an incremental cost-effectiveness ratio represents value for money. This is highly context specific, and although there are theoretical arguments for various thresholds, few countries have implemented them.

Thirdly, should a WHO guideline group wish to commission an economic evaluation for a critical recommendation the technical requirements for developing an appropriate model are substantial and are also in limited supply. Given its complexity and burden it is not generally recommended. If an economic evaluation is used, it should be transparent and comprehensive and consider sound and robust effectiveness data.

Additional reading

Eccles M, Mason J. How to develop cost-conscious guidelines. Health Technol Assess 2001;5(16):1-69. Available from: http://www.ncchta.org/fullmono/mon516.pdf

Incorporating health economics in guidelines and assessing resource impact. In: National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence. 2007. pag 66-74. Available from: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual__all_chapters.jsp

Incorporating values and preferences

From the global perspective, WHO needs to be explicit about values. 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 explicit. If differences in values are likely to affect a decision, such that people in different setting would likely make different choices about interventions or actions based on differences in their values, global recommendations should be explicit in terms of which values were applied and allow for adaptation after incorporating local values.

Additional reading

Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, Tyrer P: Systematic review of involving patients in the planning and development of health care. BMJ 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 Technol Assess 2004;8(15). Available from: http://www.ncchta.org/fullmono/mon815.pdf

Section 4: Involving patients and their representatives. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf

Formulating recommendations

A guideline panel is usually convened for a variable period (ranging from 2 to 5 days) to draft and/or review the guideline and the recommendations.

For each of the key recommendations, the quality of evidence should be specified. Panels can make a judgment about whether to grade recommendations in terms of strength; the GRADE system for grading recommendation is described below.

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 with 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 group decides on recommendations

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

It is most effective if the group considers a draft recommendation prepared by the writing team. A suggested process is listed below:

- 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, preferences
- if necessary the first recommendation is modified
- final agreement on the recommendation is reached.

If the group cannot reach consensus, one option is for the recommendation of the majority to stand, but with a note in the guideline recording 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. *Int J Qual Health Care* 2003;15:31-45

Creating guideline recommendations. In: National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence. 2007. pag 66-74. Available from: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/theguidelinesmanual2007/the_guidelines_manual__all_chapters.jsp

Dartnell J, Hemming M, Collier J, Ollenschlaeger G. Putting evidence into context: some advice for guideline writers. *Evid Based Med* 2007;12;130-2

Section 7: Forming guideline recommendations. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: 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 a recommendation outweigh the undesirable effects.
Desirable effects can include beneficial health outcomes, less burden and savings. Undesirable effects can include harms, more burden, and costs. Burdens are the demands of adhering to a recommendation that patients or caregivers (e.g. family) may dislike, 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 panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects. This can be both in favor of an intervention or against an intervention.

A **weak recommendation** is one for which the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but the panel is not confident about these trade-offs. Reasons for not being confident can include:

- absence of high quality evidence
- presence of imprecise estimates of benefits or harms
- uncertainty or variation in how different individuals value the outcomes
- small benefits
- the benefits may not be worth the costs (including the costs of implementing the recommendation)

Despite the lack of a precise threshold for going 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 below). Panels should consider all of 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. 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 bleeding risk 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 ischemic attacks	The high cost of clopidogrel and dipyridamole/aspirin as prophylaxis against stroke in patients with transient ischemic attacks

For WHO guidelines, the following options are suggested for terminology for recommendations:

strong/weak
strong/conditional
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 from: 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, Raskob G, Lewis S, Schünemann H.: Grading strength of recommendations and quality of evidence in clinical guidelines. *Chest* 2006;129:174-81

Annex 3. Methods used for developing guidelines. In: WHO Recommendations for the prevention of postpartum haemorrhage. Geneva: World Health Organization. 2007. pag 25-31. Available from: 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, Devereaux PJ, Montori VM, Schünemann HJ. Symbols were superior to numbers for presenting strength of recommendations to health care consumers: a randomized trial. *J Clin Epidemiol* 2007;60:1298-305

Research needs

Research needs and research priorities

A good guideline should summarize the available evidence as the basis of the recommendations. It is therefore also a good starting point to highlight where the available evidence may be insufficient or inadequate, or not address some of the relevant patients or public health needs. WHO guidelines should, therefore identify research needs, and if appropriate, prioritize them.

In formulating research recommendations, guideline groups should be as specific as possible about what is needed and why. A suggested format for research recommendations is 'EPICOT', described below, from Brown P et al , *BMJ* 2006 333; 804-806.

Suggested format for research recommendations 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

Peer review process

WHO guidelines should be subject to peer review, during development and prior to finalizing the draft for publication. There are several stages when peer review and external comment should be sought:

- Drafts of the questions formulated for the guideline should be circulated for comments among experts and end-users among WHO HQ, regional offices and external experts.
- Drafts of evidence profiles/tables should be circulated to experts for identification of any missing evidence, separately or in combination with -
- A pre-meeting draft of the document should be circulated widely for comment, to experts and organizations
- For full guidelines, a final draft with recommendation may be circulated for review prior to publication.

The process of reviewing comments and responding to them needs to be transparent. It is not necessary to respond to every single comment individually - and this should be made clear at the beginning of the process. However, there should be an 'audit trail' available that shows 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, you need to be clear what changes can be made. We would suggest that this be restricted to major errors of fact only.

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

- Emergency guidelines - it should be limited to review of the complete draft only, immediately prior to 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 - prior to the guideline meeting
    - a record of the response to the comments and any changes that are made
- Full - peer review would be expected to be as above, with possibly an extra review step for a second draft.

'Review by' or 'use by' date

WHO guidelines need to be issued with a 'use by' or 'review by' date - this is an indication of how long the recommendations are expected to remain valid. There is no absolute rule about how long this period may be. In deciding on the use-by date you need to take account of pace of change of research, areas where no evidence has been found and the potential need for new advice. For standard and full guidelines we suggest generally a minimum of 2 years and a maximum of 5. There also should be a description of who is the responsible person or department for initiating the review

Additional reading

Section 8: Consultation and peer review. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: 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

Updating guidelines and correcting errors. In: National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence. 2007. pag 86-9. Available from: 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, Moher D. Updating Systematic Reviews. Technical Review No. 16. AHRQ Publication No. 07-0087. Rockville, MD: Agency for Healthcare Research and Quality. 2007. Available from: 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. *Ann Intern Med* 2007;147:224-33

Section 8: Consultation and peer review. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf

Producing and publishing your guideline

You will need to 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 accurately estimate the demands that will be made on this person's time, and the impossibility of writing a major text in addition to usual duties. Once you have an idea of the approximate length of your document, you can make a rough calculation of the time needed, and commence 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

human resources. When negotiating fees and schedules, you should calculate a minimum of one week of full time work to produce 5000 words.

We strongly suggest avoiding the 'committee' approach to writing a guideline. Although getting experts to draft chapters for free may seem like a cheap and efficient way of getting the job done, unless you can guarantee quality, consistency and timely delivery, it will inevitably create more work than it avoids.

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, see link: http://intranet.who.int/homes/whp/write_edit/index.shtml.

The best editors and proofreaders are often booked up many months in advance, so make sure that you plan your 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 to be typeset. Again, WHO Press can advise on external typesetters, and the specifications that you should include when contracting for this work. GRA also provides an internal typesetting 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 templates with WHO Press before engaging an external designer. You will need a cover design, and an ISBN (international standard book number) and barcode; both of which are issued by WHO Press.

You should also give some thought to the forms in which your guideline will be disseminated, and in which format it will appear on the web. At a very minimum, you should contract your typesetter 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 your guideline, and its intended audience, you may also wish to consider providing full-text HTML, and additional materials, both electronic and printed. The WEB team is a good source of advice on improving the impact of your content.

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

Implementation and evaluation of impact

Implementation

Implementation of a guideline needs to be considered from the beginning of its development. Ideally a guideline project should be nested within a departmental or other program of work on the particular topic as 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.

There is an increasing literature on the science of implementation of guidelines and how best to encourage behavior change. The most recent Cochrane systematic review on changing physicians behaviour shows that passive educational interventions do not work, whereas small groups meetings and the opinion-leader as facilitator have a only modest impact, and audit and feed-backs techniques, incentives and mixed-strategies may have a more prominent effect. It is generally accepted at present that multiple approaches should be used, but what works best in specific settings is not clear.

The basic steps for implementing a guideline are:

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

For additional helpful reading, see Richard Grol's book, *Improving Patient Care* .

Additional reading

Grol et al. *Improving Patient Care*. Oxford: Elsevier 2005.

Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362:1225-30

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 (NHMRC). 2000. Available from: 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. *Implement Sci* 2006;1:2

Implementation of CPGs to change practice and outcomes. In: Davis D, Goldman J, Palda VA. Handbook on clinical practice guidelines/Canadian Medical Association. Ottawa: Canadian Medical Association. 2007. pag 17-29. Available from:

http://www.cma.ca//multimedia/CMA/Content_Images/ClinicalResources/PDF/English/CPGHandbook.pdf

Section 10: Implementation. In: SIGN 50: a guideline developer's handbook. Revised edition 2008. Edinburgh: Scottish Intercollegiate Guidelines Network. 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf

Evaluation / monitoring

The guideline should define parameters or outcome measures that can be monitored for the main recommendations. Ideally, there should be measures at baseline to define 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 who is the target audience for the guideline and the plans for implementation.

Additional reading

H M Hearnshaw, R M Harker, F M Cheater, R H Baker, G M Grimshaw. Are audits wasting resources by measuring the wrong things? A survey of methods used to select audit review criteria. *Qual Saf Health Care* 2003;12: 24-28