Swiss TPH
Swiss Tropical and Public Health Institute
Schweizerisches Tropen- und Public Health-Institut
Institut Tropical et de Santé Publique Suisse

Associated Institute of the University of Basel

Swiss Centre for International Health

Handbook for Supporting the Development of Health System Guidance

Supporting Informed Judgements for Health System Policies

Xavier Bosch-Capblanch
On behalf of the project team
Swiss Tropical and Public Health Institute
Associated Institute of the University of Basel

July 2011

The Swiss Centre for International Health is a World Health Organization Collaborating Centre for Health Systems Development

Swiss TPH
Swiss Tropical and Public Health Institute
Schweizerisches Tropen- und Public Health-Institut
Institut Tropical et de Santé Publique Suisse

Swiss Centre for International Health
Swiss Tropical and Public Health Institute
Socinstrasse 57
P.O.Box
4002 Basel
Switzerland
Internet: www.swisstph.ch

Contact

Dr. Xavier Bosch-Capblanch. SCIH, Swiss Tropical and Public Health Institute

Deputy Head Systems Performance and Monitoring Unit

Socinstrasse 57, PO Box 4002 Basel (Switzerland)

Telephone direct line: +41 61 284 83 19; Fax +41 61 284 81 03
E-mail x.bosch@unibas.ch

Website http://www.swisstph.ch/

Contributors

Project team	
Xavier Bosch-Capblanch (project leader)	Swiss Tropical and Public Health Institute
Lise Beck	Swiss Tropical and Public Health Institute
Daniel Dröschel	Swiss Tropical and Public Health Institute
Simon Lewin	Norwegian Knowledge Centre for Health Services (Norway)
Peter Steinmann	Swiss Tropical and Public Health Institute
Svenja Weiss	Swiss Tropical and Public Health Institute
WHO partners	
Tikki Pang	WHO/HQ
Abdul Ghaffar	WHO/HQ
Regina Kulier	WHO/HQ
Metin Gulmezoglu	WHO/HQ
Davina Ghersi	WHO/HQ
Ulysses Panisset	WHO/HQ
Carmen Dolea	WHO/HQ
Timothy Evans	WHO/HQ
Task Force	
Andy Haines (chair)	London School of Hygiene and Tropical Medicine (UK)
Edgardo Abalos	Centro Rosarino de Estudios Perinatales (CREP) (Argentina)
Rifat Atun	The Global Fund to Fight AIDS, Tuberculosis and Malaria (Switzerland) and Imperial College, London /UK)
Fadi El-Jardali	American University of Beirut, AUB (Lebanon)
Lucy Gilson	University Cape Town (South Africa), LSHTM (UK)
John Lavis	McMaster University (Canada)
Sandy Oliver	Institute of Education, University of London (UK)
Pierre Ongolo-Zogo	University of Yaoundé (Cameroon)
Jon-Arne Rottingen	Alliance for Health Policy and Systems Research (Norway)
Peter Tugwell	McMaster University (Canada)
Gunn Vist	Norwegian Knowledge Centre for Health Services (Norway)

Acknowledgements

- ◊ Thanks to the Task Force for many useful suggestions, comments and follow up of project activities. To Edgardo Abalos for providing sources for health systems guidance (Trip database); thanks to Gun Elisabeth Vist for the document containing references of potential sources for health systems guidance (Håndbok for Nasjonalt kunnskapssenter for helsetjenesten – 2009).
- ◊ Thanks to Roser Cussó, Associate Professor Universitat Politècnica de Catalunya, for discussions on systems frameworks; to Don de Savigny for discussions on knowledge translation topics.
- ◊ Thanks to Katharina Schwan, intern at the Swiss Centre for International Health (Swiss TPH) for proof-reading the first version of the handbook.
- ◊ Thanks to the WHO staff participating in the lunchtime seminar, 1st March 2010.
- ◊ Thanks to those attending the presentation on the development of the handbook at the First Global Symposium on Health Systems Research, 18th November 2010, for their questions and comments.
- ◊ Thanks to the Rockefeller Foundation for the financial support.

Disclaimer

The views and ideas expressed herein are those of the author(s) and do not necessarily imply or reflect the opinion of the Institute.

Abbreviations

COGS	Conference on Guideline Standardisation
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GRC	Guidelines Review Committee
LHW	Lay Health Worker
M&E	Monitoring and evaluation
PICOT	Participants – Intervention – Comparator – Outcome – Time.
SCIH	Swiss Centre for International Health
Swiss TPH	Swiss Tropical and Public Health Institute
WHO	World Health Organisation

Table of contents (abridged)

Table of contents (detailed)

Annexes

List of tables

List of figures

About the handbook

This handbook, commissioned by the WHO, describes the processes, approaches and outputs for developing health system guidance and is compliant with the existing 'WHO handbook for guideline development'[1] (WHO Guidelines Review Committee (GRC))[1] and is the equivalent of the handbook to support the development of clinical guidelines for health systems guidance. It is based on a preliminary work that established the rationale and framework for health systems guidance (see Figure A - 1)[2] and it is inspired by global trends encouraging to bridge the gap between research and policy and practice through knowledge translation (see Annex 3). The handbook has been produced by a core team supported by the GRC staff, supported by a Task Force specifically set up for this project. The handbook deals with the process of developing full guidance, rather than the processes to adopt, adapt or endorse guidance developed by third parties.

'Guidance' is defined as the "systematically developed body of knowledge, integrating research evidence and descriptions of the types of other considerations needed to inform decision making about appropriate health system arrangements in specific settings". We have used the term 'guidance' as distinct from guidelines to clarify that the process of supporting evidence informed judgements for health systems issues is different in a number of respects from the process of developing clinical guidelines (e.g. strong influence of contextual factors on appropriate decisions or the range of research designs used to develop evidence)[2].

'Guidance' is understood as the whole body of knowledge that informs policy decisions on health systems arrangements. 'Guidance' is understood as a 'policy-oriented' product (as opposed to a 'research product') ready to be used by global and national policy making initiatives as well as by health care providers, the population affected by health systems policies and other stakeholders Guidance bridges (translates) research synthesis (and other types of evidence) and policy needs for evidence. Health systems guidance supports decisions about how to organize health systems and how to get effective programs, services and drugs to those who need them; not about which public health programs or drugs to fund or deliver.

The handbook is structured in chapters which follow the chronology of guidance development (see Figure 1). It contains:

- background information explaining the rationale for the approaches proposed to develop guidance. This should allow guidance developers to consider alternative approaches that can better suit the needs of specific guidance developments;
- methods and outputs for each phase of guidance development; these can range from full descriptions to references to methodological descriptions, depending on how well methods are developed, validated or widely accepted. At this relatively early stage in the development of methods to produce health systems guidance, it would be difficult and may be reductionist to offer 'cookbook' approaches;
- topics that require further research (see Annex 13);
- references to the WHO guideline handbook (marked in the margin with 'GRC');
- additional information (mostly in the Annexes), examples, and forms.

[1] Compliance with the GRC procedures only applies to guidance produced within WHO.

[2] The term 'guidelines' is still used when referring to clinical guidelines or when citing references that use this term.

The ‘handbook’ is conceived as an iterative process that will need further developments to address the conceptual and methodological challenges of developing health systems guidance and it is meant to be piloted by guidance developers at WHO to inform a second edition of the handbook. Alongside the research topics suggested in the handbook, further activities needed to keep it up-to-date may include:

- actively disseminate the handbook among potential users in other institutions, potential users of guidance and researchers;
- promote the participation and the adoption of the handbook by stakeholders involved in health systems guidance production;
- develop a web based interactive version of the handbook using multimedia resources, including a learning programme on (health systems) guidance development[3,4];
- document handbook utilisation and enrich it with case studies;
- increasingly incorporate needs and views of key stakeholders;
- monitor new advancements in relevant methods to guidance development, including the incorporation of qualitative research evidence, colloquial evidence, Monitoring and Evaluation, policy analyses and health systems thinking;
- establish a research agenda around the approaches and methods to develop, disseminate and implement health systems guidance; formally liaise with ongoing efforts in the fields of research synthesis (e.g. Cochrane and Campbell collaborations), evidence appraisal (e.g. Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, AMSTAR) and guidance development (e.g. AGREE Collaboration, international guidelines standards development);
- ensure long term funding.

1 Introduction: health systems guidance development processes

When you set sail for Ithaca,
wish for the road to be long,
full of adventures, full of knowledge.
(Constantine P. Cavafy. Ithaca)

There is a great amount of research, research synthesis and guidelines on specific diseases. However, despite the availability of interventions that address the major health problems of the population, weak health systems[5,6] and the lack of appropriate policies, hamper equitable access to them[7]. Hence, the renewed efforts to address health systems issues under the term of 'health systems strengthening[8,9] as well as the growing number of WHO Member States and political and international health leaders who recognise the need to make a major, sustained commitment to strengthening health systems" [10].

Decisions to address health systems issues must be based on the best available evidence. Guidance is a transparent way of producing recommendations on the possible courses of action to address a problem. At present, however, there is no accepted international standard for guidance development[11].

Research topic 1. Methods for developing health systems guidance.

There are several handbooks and approaches to produce guidance. Most of them share the same methods. However, there is no accepted 'best practice' on how to develop guidance and even less clarity on the particularities of guidance development for health systems.

This chapter presents an overview of the guidance development process, pointing at the specific issues in developing guidance for health systems. It also contains a brief description of the resources needed.

1.1 Generic guidance development processes

Any publication containing WHO recommendations has to be approved by the WHO's Guidelines Review Committee (GRC)[1], regardless of the type of document. Other institutions may have similar review bodies. The term guidelines (and by extension 'guidance'), in the WHO context, applies to any document containing a recommendation on a course of action. This handbook complies with the WHO's GRC procedures and recommendations.

The GRC is the body, at WHO, that has the mandate to review and approve guidelines. All guidelines follow the same process (Figure A - 2) and are reviewed twice by the GRC: once after scoping the guidelines ('initial approval for development', see Figure A - 3, Annex 4) and once after the recommendations have been issued ('final approval', see Figure A - 4, Annex 4).

WHO distinguishes several types of guidance: 'rapid advice', 'standard', 'full', 'compilation of' and 'adaptation of' guidance. In essence, rapid advice guidance is considered when an urgent issue needs guidance and there is no time to gather all the available evidence on the topic. Standard and full guidance follow the procedures and systematic methods in the development of guidance. 'Standard' has a narrower scope that focuses on specific interventions, as opposed to 'full' guidance which covers all aspects of a topic. When guidance exists, this can be either updated with new recommendations or existing recommendations can be compiled into a single document[1]. This handbook deals with 'standard' and 'full' guidance, with no distinction between the two. Rapid

methods, adaptation and compilation of guidance may need other approaches and methods than those described in this handbook. All types of guidance go through GRC approval.

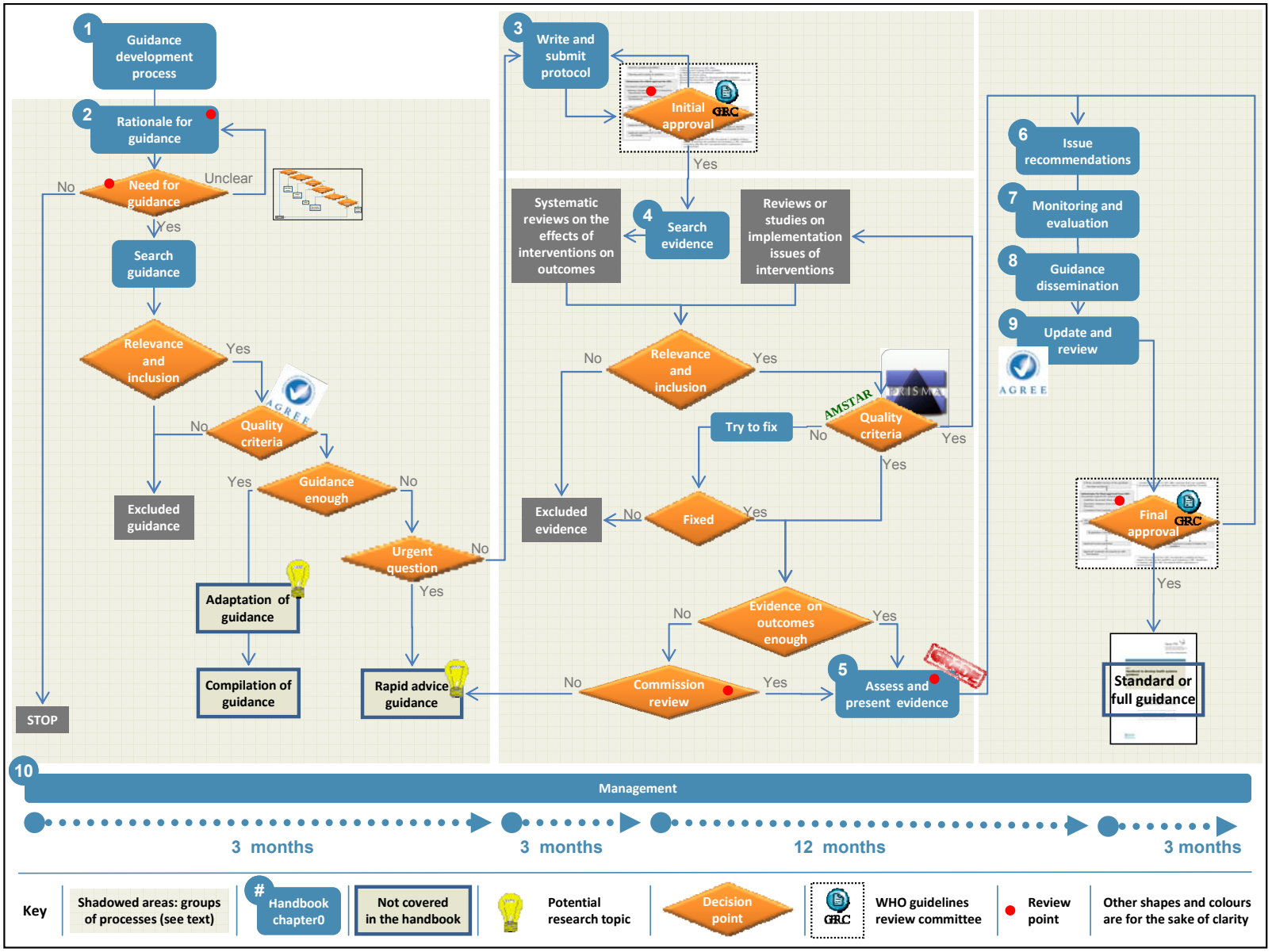
Guidance development involves the following groups of processes (see big shadowed areas in Figure 1):

1. establishing the need for guidance: whether the issue of interest should be addressed with guidance and, in that case, whether new guidance has to be developed or existing guidance can be adapted and used (chapter 2);
2. preparations to develop health systems guidance, including the protocol that details the methods to be followed (chapter 3);
3. retrieval and appraisal of evidence to support guidance (chapters 4 and 5);
4. issuing, presenting and disseminating recommendations (chapters 6 to 9).

Each process involves several steps. Figure 1 describes a detailed flow chart of the development of guidance, with the key decision points. The figure shows:

- the key decisions to be taken in the process of guidance development;
- the initial and final approval by the GRC;
- the main steps and tools used;
- the different types of guidance products, as mentioned above;
- the corresponding chapter numbers in this handbook.

Figure 1. Guidance development flowchart.



The processes start with the guidance steering group familiarising itself with the guidance development process, including regulations, administrative and formal requirements. The overview of the processes will assist guidance developers in having an initial view of requirements in terms of resources and time to develop guidance. Details on the time frame and workload can be found in Chapter 10.

Next, guidance developers have to define the rational to develop guidance; essentially: to which extent is it justified to develop guidance. This requires an initial scoping, during which issues such as demand for guidance (e.g. who asked for it, what is the desirable timing), audience (e.g. policy makers, managers), type of problem to address (e.g. health system, a particular problem or a systemic problem), priority issues (i.e. in relation to other guidance, in relation to other problems) or level of guidance needed (e.g. global, regional) are considered. Full scoping of guidance is done in Chapter 2.

Then, before engaging in the actual production of guidance, it is essential to verify whether existing guidance can be used or adapted. Based on the scope defined earlier, guidance is systematically searched and included if it meets the inclusion criteria and level of quality. If appropriate guidance of reasonable quality still exists, it may be used or adapted as needed. If new guidance needs to be produced it is important to establish a protocol that details all steps for guidance development. The details of the protocol will span from sources and strategies to searching evidence, concluding with strategies for the dissemination of guidance and maintenance to keep it current. (see Chapter 3). This protocol needs to be approved by the GRC ('initial approval').

Once the protocol is approved, the work can continue by searching the evidence according to the scope of guidance. The preferred sources of evidence are systematic reviews or overviews. However primary research may have to be used in the absence of systematic reviews that comply with the full scope of guidance. Evidence on the effects on outcomes and implementation issues is searched. Reviews are assessed for their relevance, inclusion and quality*. Some quality issues can be fixed (e.g. contacting the authors of the reviews) and some evidence may be discarded due to the the low level of quality (Chapter 4). If available evidence is judged sufficient to develop guidance, the steps described in this handbook are followed. However, if available evidence is not sufficient, the guidance steering group may consider commissioning a systematic review of missing evidence or producing rapid advice guidance instead. If evidence is enough, this evidence is appraised by using widely accepted tools (e.g. Grading of Recommendations Assessment, Development and Evaluation (GRADE)) and properly presented to the group responsible for issuing recommendations (Chapter 5). The evidence is critically reviewed to ensure that complies with the scope of guidance and that it is of reasonable quality to inform recommendations. Insufficient evidence may require additional searches. Once this step has been cleared, there should be no need to revisit the quantity and quality of evidence in later stages.

Recommendations are issued based on the evidence on outcomes and on implementation issues. A deliberative process is set up to reach a common understanding and consensus among researches, policy makers, potential users of guidance and other relevant stakeholders (Chapter 6). Guidance also has to provide clues on how to monitor and evaluate the recommendations issued. This is particularly important in health system guidance, where evidence on health system-level interventions tends to be weaker, and implementation issues more relevant. Guidance on monitoring and evaluation is addressed in Chapter 7. The mechanisms to ensure effective

* 'Quality of reviews' describes to what extent reviews have been conducted and reported using acceptable standards. A later issue is the quality of evidence, i.e. the degree of confidence that can be placed in the research findings (which are reported in reviews). Although a systematic review can be done complying with the best standards, the evidence used in the review may be of poor quality due to limitations in the primary research.

dissemination and use of guidance are addressed in Chapter 7.2.3. This chapter will include suggestions on how to monitor guidance use.

At the end of the guidance development process, a formal review of the completed work takes place. Guidance is assessed against accepted guidance quality criteria (e.g. AGREE) and an explicit plan to address pending issues (e.g. missing evidence) and to update guidance is created (Chapter 9). The completed guidance must be submitted to the GRC for final approval before it can be institutionally endorsed and distributed.

As can be seen in Figure 1 (red bullets), there are several critical steps which need critical review by the GRC and/or external reviewers:

- review and approval of the protocol to develop guidance (GRC);
- review and approval of the final body of evidence that will support the recommendations;
- final review and approval of guidance (GRC).

1.2 Health systems guidance development processes

The processes, approaches and tools in this handbook deal with the challenges of producing guidance for health system-level interventions. There are several adaptations from the existing WHO generic guidance development process that account for the specific aspects of research evidence on health systems and the somehow more complex implementation issues when addressing health systems issues. Table 1 describes in parallel the main steps in those processes.

The main steps where adaptation is needed are the following:

- it is proposed that a full protocol for guidance development is submitted for approval, including review by an external review group. This does not substitute the current submission forms to the GRC, but is meant to improve the planning process, to anticipate bottlenecks and to ensure transparency and objectivity;
- the rationale for guidance development on health systems does not necessarily follow a single research question (i.e. policies may refer to a set of services or to certain components of the health systems, while research is typically done on a single or a limited number of interventions);
- the evidence on outcomes is as relevant as the evidence on implementation issues of the intervention itself and on other parts of the health system that may be affected by the intervention(s) of interest; as well as a wider scope of contextual issues that need to be systematised;
- the tools to assess the quality of evidence are based on GRADE but will need some adaptation or additional items regarding certain study designs used in health systems research;
- recommendations on options, including pros and cross of each interventions, are also based on a wider range of considerations around health systems and implementation issues;
- suggestions on the evaluation of recommendations are considered prior to the dissemination of guidance, as they are linked to each specific recommendation.

Table 1. WHO generic and health systems guidance development processes.

Generic WHO guidance development	Handbook to develop health systems guidance
1. Setting up WHO Steering group and scoping of the document	No change (see **Chapter 10**)
2. Setting up Guideline Development Group and External Review Group	No change (see **Chapter 10**)
Initial guideline approval	
3. Management of Conflicts of Interest	No change
4. Formulation of the questions and choice of the relevant outcomes	**Chapter 2** (rationale) and **3** (protocol) It is proposed that the initial guideline approval be based on a protocol for guidance development which includes the formulation of questions and the methods to be used.
	Initial guideline approval
5. Evidence retrieval, assessment and synthesis (systematic review(s)). GRADE - evidence profile	**Chapter 4** (retrieve evidence) and **5** (assess and present evidence) It is proposed that evidence is retrieve on (1) outcomes and (2) implementation issues. The quality of evidence may require additional items which are not contemplated in GRADE. Also issues about how to present evidence are addressed.
6. Formulation of the recommendations (GRADE). Including explicit consideration of: benefits and harms, values and preferences, resource use	**Chapter 6** (recommendations) It is proposed that recommendations on options are based on (1) to what extent the intervention(s) work(s) and (2) to what extent they can be implemented. They are complemented with pros and cons. Additional criteria from those included in GRADE (apart from benefit, harms, values, preferences and resources) may need to be used.
Final guideline approval	
7. Dissemination, implementation (adaptation)	**Chapter 7.2.3** (dissemination)
8. Evaluation of impact*	**Chapter 7** (evaluation)
9. Plan for updating	**Chapter 9** (appraising and updating)
	Final guideline approval
	Chapter 10 (management)

These adaptations are further detailed in each chapter.

1.3 Management considerations for guidance development

It is useful to have a basic understanding of the burden of developing health system guidance from the very beginning. This is further explained in the chapter dealing with managerial aspects (Chapter 10).

* Not yet fully developed in the existing handbook (see page 61)[1].

People

A number of different people involved in the processes to develop guidance are able to fulfil the diverse functions and technical expertise needed.

A guidance steering group will manage the whole processes and ensure adherence to best practices. The group is lead by the coordinator who has the overall responsibility while group members engage in technical tasks. The guidance development group involves a wider range of constituencies who assist the steering group in supervising the quality of guidance and making it relevant (e.g. researchers, policy makers, potential users and civil society representatives). A similar group of panellists is responsible for deliberating on the available evidence and to issue recommendations accordingly.

In the WHO generic guidelines handbook[1], the guidance development group is also in charge of developing recommendations. However, the deliberative process for developing recommendations needs people with profiles that may be different from those needed in the guidance development group. The deliberative process (as further described in Chapter 6) requires a balanced mix of expertise, a certain number of members, communication skills and the capacity to engage in dialogue. Therefore it is suggested to have an explicitly different group of panellists to issue recommendations.

The tasks involved in developing guidance need external technical assistance, in areas such as literature searches, interpretation of statistical outputs, design of dissemination materials or translation. The people involved in guidance development are complemented with external reviewers who will independently provide advice and criticism in critical phases of the guidance development processes. See Table 42 (page 120) for an overview of these groups. See chapter 10 for explanations on people management, including conflicts of interest. In summary:

- Guidance steering group and coordinator: manage and drive the guidance development process and undertake routine activities.
- Guidance development group: assists the guidance steering group and reviews the guidance outputs at several stages of the process.
- Guidance panel for recommendations: gathers a wide range of stakeholders to issue recommendations.
- External reviewers: review guidance outputs at several stages of the process.
- Technical support:
    - Literature searches specialist
    - Researchers, statisticians
    - Communication experts and workshop facilitator
    - Translators
    - Other

Other resources

The development of guidance needs two types of support materials: templates and tools:

- templates are structured documents which need to be filled to record the steps followed in guidance development (e.g. conflict of interest forms, evidence profiles);
- tools are methodological instructions that are followed in several steps of the guidance development processes (e.g. GRADE system to assess the quality of evidence)

Guidance documents (e.g. protocol, final guidance document), tools and templates are stored using a standard filling system.

At the bottom of Figure 1, there is also a timeline showing the approximate duration of the main processes. Systematic reviews can take typically between one and one and a half years and

guidance development can take one year and a half or more[12]. Development of standard guidance is a lengthy process which needs to keep people engaged. It is estimated that a reasonable budget for the full process is around 100,000 USD[1], including the costs for certain external technical assistance and excluding conducting systematic reviews. All these aspects are fully developed in Chapter 10.

2 How to establish the rationale for developing health systems guidance

"Do cats eat bats?" […] "Do bats eat cats?"
for, you see, as she couldn't answer either question,
it didn't matter which way she put it.
(Lewis Carrol. Alice adventures in Wonderland)

This chapter addresses the initial step in guidance development: establishing the rationale for guidance. First: frame the problem for which guidance would be developed. Second: decide whether developing guidance is necessary by using a systematic screening process which takes into account already existing guidance, among other issues.

2.1 Framing the 'problem' for health systems guidance

Problems addressed by health systems guidance are health systems problems: 'health systems' are the problem[13]. The health system problem may have been identified by policy makers, users of health services or other stakeholders, and brought to the attention of an institution able to produce guidance.

Ideally, the 'problem' is what triggers a policymaking process[14]. Focusing on health systems problems should not lead to forgetting that people are at the centre of health care[15] and that people's needs should be reflected[14] in policy formulation as well as in research synthesis[16] and primary research[17]. In other words, health systems guidance addresses health systems problems but the perspective of populations affected by how the health systems works has to be explicit. Users' centeredness has to be explicit in the problem definition.

Defining the problem is at the core of guidance development because it will determine all the following steps in guidance development[1,17]. There does not seem to be a comprehensive and validated approach for problem definition. We suggest some approaches[17] and additional resources from the 'quality assurance' community to define and model problems.[18].

Table 2 frames the problem in terms of the health related topic it is associated with, its parameters, a modelling exercise to establish plausible causal links, the context and opportunities relevant for guidance development, in addition to some suggested approaches and an example. The last column in the table highlights how framing the problem will determine the next steps in the guidance development processes. The implications for evidence and guidance (last column in Table 2) show how framing the problem will determine the needs for evidence, the focus of guidance and, eventually the focus of policy. Accuracy at this stage is essential. For example, modelling the problem will inform which intervention or set of interventions should be considered in the guidance development process, and provide initial insights on how they can be articulated into meaningful advice for policy makers.

The current handbook for guideline development[1] addresses the framing of the problem in terms of a 'formulation of questions' to be addressed by guidance after the initial approval process. This can be suitable for clinical or public health guidance as it is easier to frame and more prone to be addressed with a single intervention. However, in health systems guidance, defining and modelling the problem may be a more complex task needing several iterations to identify which 'group' of problem-factors need to be addressed by guidance. This will also define the search for existing guidance and the rationale for developing new guidance. Hence, when considering the development of health systems guidance, problem definition should take place at this stage. The current

handbook suggests prioritising the ‘questions’; while in health system guidance it may be more appropriate to ‘frame and model the problems’ rather than individual questions per se[*].

[*] The problem framed will be formulated in terms of PICOT questions to guide the search of evidence, but it is useful to start with a properly framed problem, which is what, ultimately, guidance will address.

Table 2. Framing the problem for health system guidance and implications for inclusions of evidence and guidance.

	Description	Suggested approaches	Examples	Implications for inclusion of evidence and guidance
Main topic	Definition of the area or discipline the problem belongs to	Use existing frameworks and initiatives	Human resources for health[19]	Scope (see below for details)
Problem statement	Clear and brief statement of the problem	Reflect users' centeredness and health system issue	People worldwide in rural and remote areas have less access to trained health workers ("*equity to meet the health needs*")	(See below, in 'Problem parameters')
Problem parameters	Determinants of health and health status	Basic demographic profile	"*people living in rural and remote locations*"	(Users) participants
	In terms of health system	Rural / urban; primary, secondary, tertiary; national, provincial, districtal	"*attract and retain health workers in rural and remote areas*"	(Health system) participants
	Population and health systems events by which the problem is identified as such	Quantification of the magnitude and severity for main events Consider cross-cutting issues (e.g. equity or gender)	"*access to trained health workers*"	Outcomes and evaluation parameters
Problem modelling	Establish a causal chain linking health systems problems with users' needs	Use of accepted frameworks; consider the 'role' of different constituencies; incorporate parameters described above.	"*Insufficient numbers and types of qualified health workers in remote and rural areas impedes access to health-care services for a significant percentage of the population*"	(Groups of) interventions
Context	Geographical context, stakeholders and other factors influencing people and health systems.	Existing classifications of countries (e.g. by income); formal stakeholders analyses	"*Policy-makers in all countries, regardless of their level of economic development*"	Scope of guidance (e.g. audience)
Opportunities	Establish what makes your institution at the present time suitable to address the problem	Visibility of the problem in the international community; availability of funds for the topic; new and relevant evidence available; local policy 'momentum'.	"*In June 2009 the high-level Taskforce on Innovative International Financing for Health urged all governments to ensure that all people, including rural and remote populations, have access to safe, high-quality and essential health-care services*"	Audience of guidance

Note: parts in italics are literal extracts from existing guidance on rural retention of health workers[20]. See also Table 34. Framing the problem for health system guidance and implications for monitoring and evaluation.

2.1.1 Approach: framing the problem, parameters and models

a. Modelling the problem

Scoping the problem for health systems guidance can be an endless task in itself. It is proposed that this is done iteratively by defining the parameters and revisiting them as needed. It may be useful to start by modelling the problem in a graphical way to identify plausible causal relations to target the problems that may be the subject of guidance.

Based on the WHO health system building blocks framework[21] (Figure 2), the example in Figure 3 shows plausible causal relations in the case of 'pay for performance'[22]: in this example the relations between health system components are expressed with arrows suggesting positive and negative interactions ('+' and '-' signs). Additional factors that affect these relations are also included close to the arrows.

Figure 2. WHO health systems framework.

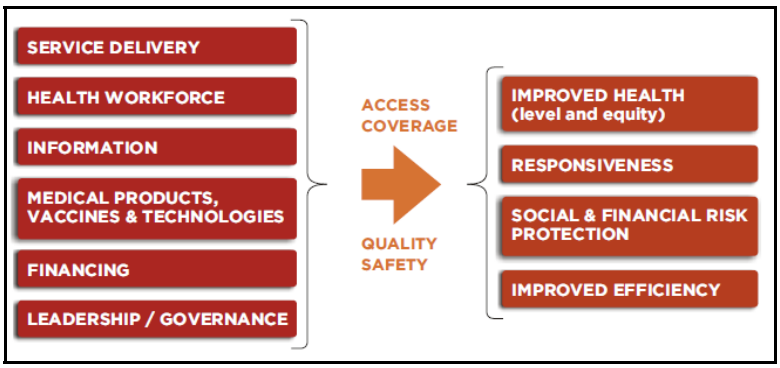
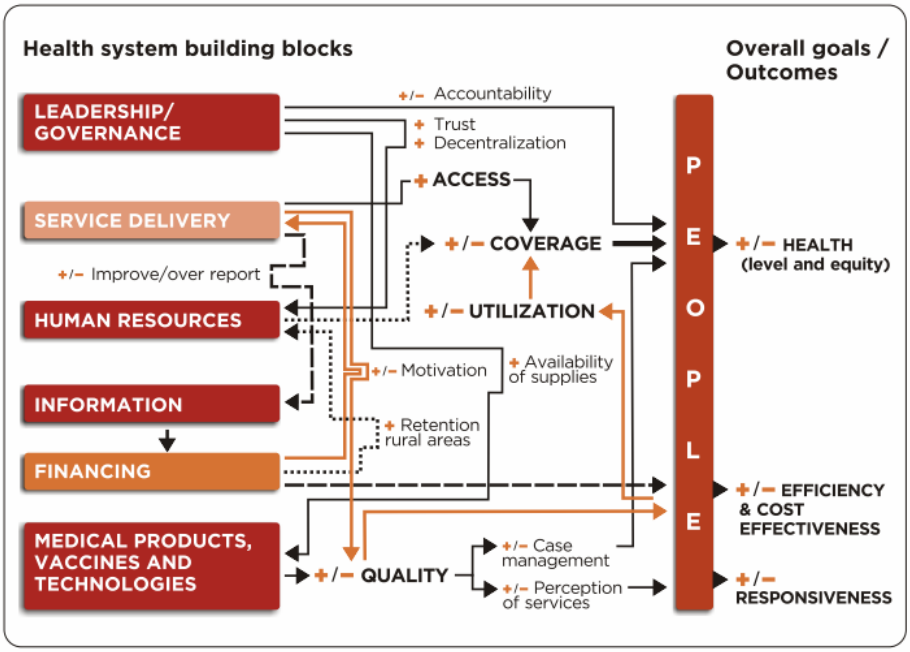


Figure 3. Conceptual pathway for the 'pay for performance' intervention.



The definition of parameters is not a neutral exercise, but it entails value judgements; e.g. rural-urban differentials in the access to health care could be made either explicit or overlooked. Even if these differences are quantified, they could be seen as either

irrelevant or as crucial. The categories ‘determinants of health’, ‘health system’ and ‘health status’ (see Table 2) can be useful to remember that ‘users’ (people in the Figure) and ‘health systems’ perspectives have to be taken into account.

b. Measuring the problem

Problem parameters can be described qualitatively and quantitatively. They can also assist in identifying the indicators that describe the problems and the sources of information[23], which come from the specific countries situations. Information on these categories is expected to be increasingly available as information systems improve to provide data for countries’ minimum set of health indicators[23].

The Health Metrics Network defines several data sources[23] which can be of value to describe the problems:

- Censuses: size of a population, its geographical distribution, social, demographic and economic characteristics of people.
- Civil registration: the continuous recording of vital events (live births, deaths, foetal deaths, marriages and divorces) in accordance with the legal requirements in each country.
- Surveys for child and maternal mortality and health, nutrition, service use, and knowledge and practices related to health care; health status evaluations, descriptions and determinants; knowledge, beliefs and practices related to disease prevention and transmission; household expenditures on health; and inequalities in health outcomes and access to health services; surveys on health systems issues, such as workforce surveys.
- Individual records: documentation of the provision of health services (for example, of growth monitoring, or antenatal and delivery of care) to individual clients; case reports and disease records routinely produced by health workers; and information held in special disease registries (such as those for cancer).
- Service records: health service providers and records of events with important health consequences produced in other sectors (police, veterinary services, environmental health authorities, insurance companies and occupational health agencies). They include unintentional injuries, homicides, suicides, road traffic accidents, environmental and meteorological incidents and alerts on food and product safety. Includes the private sector as well.
- Resources records: resource and administrative records focus on the quality, availability and logistics of health service inputs. This includes data on the density and distribution of health facilities, human resources for health, budgets and expenditures, drugs and other core commodities, and key services.
- Research: biomedical, public health, health policy and systems, environmental health, social and behavioural sciences, operations research.

Table 3 suggests the most appropriate data sources for each category of the problem definition to assist the guidance steering group in focusing on the most relevant information.

Table 3. Data sources relevance for categories of the problem definition.

	Determinants of Health	Health systems		Health status
		Inputs and outputs	Outcomes	
Censuses	☑	☑		☑
Civil registration	☑			☑
Population surveys	☑	☑	☑	☑
Individual records	☑	☑	☑	☑
Service reports		☑	☑	☑
Resource records		☑		
Research*	☑	☑	☑	☑

Note: adapted form the Health Metrics Network[23].

The indicators are specific to the parameters that define the problem. This is essential to ensure that an effort was made to qualify or quantify what was considered important when framing the problem. If data is not available for any parameter of the problem or any item in the model or framework, this needs to be made explicit. Disaggregated data may also help to unveil a problem or part of a problem that was overlooked. Indicators have to be interpreted in a comparative manner to give a more accurate profile of the problem[17], such as comparing geographical regions in time or against accepted benchmarks or targets. Research can be a useful source of information to establish comparisons.

Identifying indicators and searching, processing and presenting evidence on the problem(s) can be very demanding tasks. Hopefully, most of the information needed to define a problem is already organised and digested in articles, reports or statistical summaries. It is also expected that data, information and reports on problems that are considered a priority tend to provide overwhelming evidence on the magnitude of those problems. This is a task that has to be kept in scale because it would not be possible to gather all possible data on any single problem.

The quality of local data, which constitutes the source of evidence informing the parameters of the problem, has to be appraised as well[24]. The Data Quality Assessment Framework[25] can be a useful source of items to be considered when assessing the quality of data. The following table describes some of the items[23]:

* Includes qualitative research.

Table 4. Criteria to assess the quality of data.

Item	Description
Methodological soundness	The methodological basis for the statistics follows internationally accepted standards, guidelines, or good practices
Accuracy	Source data and statistical techniques are sound and statistical outputs sufficiently portray reality
Timeliness	The period between data collection and its availability to a higher level, or its publication.
Periodicity	The frequency with which an indicator is measured.
Consistency	The internal consistency of data within a dataset as well as consistency between datasets and over time; and the extent to which revisions follow a regular, well established and transparent schedule and process.
Representativeness	The extent to which data adequately represent the population and relevant subpopulations.
Disaggregation	The availability of statistics stratified by sex, age, socioeconomic status, major geographical or administrative region and ethnicity, as appropriate.
Confidentiality	Data security and data accessibility – the extent to which practices are in accordance with guidelines and other established standards for storage, backup, transport of information (especially over the Internet) and retrieval.

Note: adapted from the Data Quality Assessment Framework[25].

c. Context and opportunities

The way a problem is framed has to take into account contextual issues as well as opportunities that may facilitate or block the guidance development processes (e.g. availability of funds, international agendas). There must be a consensus between different groups because framing a problem is not a completely neutral and objective task (e.g. policy makers, population ultimately affected by the problem, stakeholders). Neither users, nor policy makers are homogeneous groups of people; some may be more sensitive to problems linked to specific diseases, access to services or financial aspects. It is important that the key constituencies are brought on board from the start of guidance development. Even when consensus is not possible, it is important to understand why it is not.

2.1.2 Outputs: problem framed

The framed problem serves two purposes: to assist in the process of identifying existing guidance (and then decide whether new guidance is needed) and to inform new guidance development, if it really needs to be developed. As mentioned above, this is a crucial step in shaping how and what to do next. Therefore, it is proposed to submit the problem framed for review in a background document (for an example, see the background document for the rural workforce retention guidance[26]).

There are several options on how to structure the background document. In general:

- it should contain clear definitions of the terms used;
- it should have a background placing the problem in the wider context of international health and health systems;
- it is centred in the description of the problem, in a systematic way (following Table 2);
- it can include considerations on alternative ways of framing the problem;
- it ensures that key cross-cutting issues (e.g. equity, gender) are included, where relevant;
- it shows indicators (qualitative and quantitative, as needed) and makes explicit those indicators for which no data or information is available;
- it provides and appraises evidence that supports the descriptions of the parameters of the problem and the model;
- it makes explicit any assumptions and sources of evidence;
- it includes preliminary hints on the availability of evidence and guidance to address the problem, if necessary;
- it is written using plain language where possible, or includes a plain language summary (one page);
- it adheres to some or all items of reporting standards (e.g. of observational studies[27]);
- it is brief.

Figure 4 is an example on how to present a one page summary of a framed problem using the example of rural retention of health workers[20].

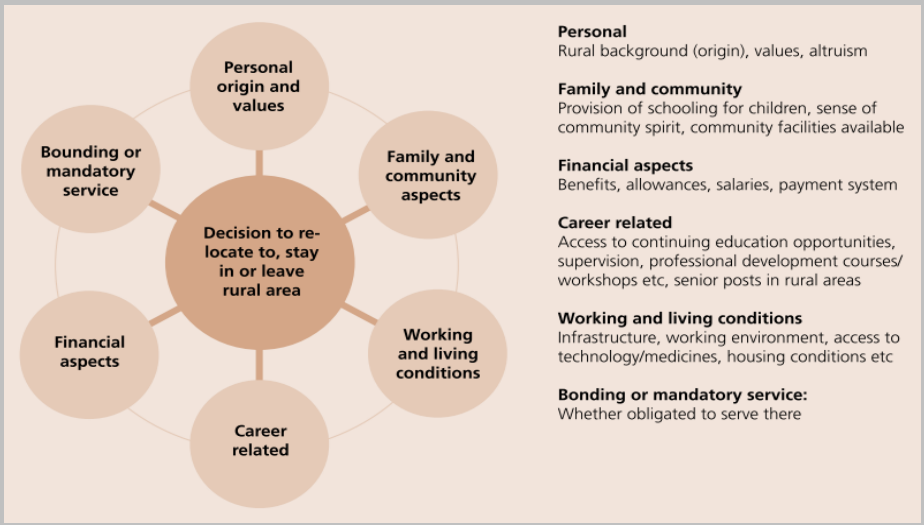
Figure 4. Example of problem framing.

PROBLEM FRAMING

Topic | **Human resources for health**

Problem statement	People worldwide in rural and remote areas have less access to trained health workers
Problem parameters	
• **Determinants of health**	People living in rural and remote locations Less access to health services for a significant proportion of people
• **Health status**	-
• **Health systems**	Health system in rural areas
○ **Human resources**	Insufficient numbers and types of qualified health workers in remote rural areas.
○	
Context	All countries, regardless of their level of economic development
Opportunities	Taskforce on Innovative International Financing for Health (2009) urged all governments to ensure that all people, including rural and remote populations, have access to safe, high-quality and and essential health-care services

Problem model or framework:



Indicators	Value	Place	Year	Reference
Rural population	50%	World		
Urban population	50%	World		
Rural nurses	38%	World		
Urban nurses	62%	World		
Rural physicians	24%	World		
Urban physicians	76%	World		
Rural population	20%	USA		
Urban population	80%	USA		
Rural physicians	9%	USA		
Urban physicians	91%	USA		

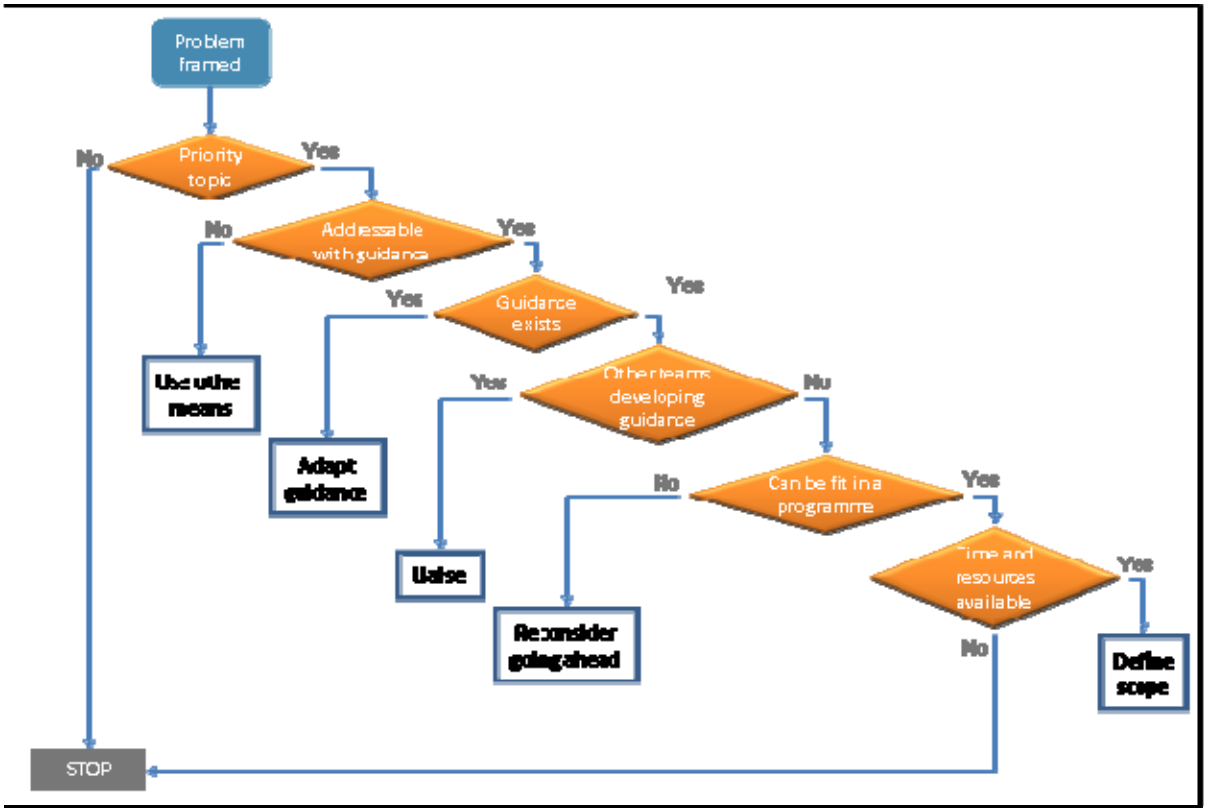
(See background document)

2.2 Establishing the need for health systems guidance

2.2.1 Approach: screening process

The process to establish the rationale for developing health systems guidance needs to be explicit and transparent. A decision flowchart is proposed to assist this process. Once the problem has been framed (see preceding section), a series of questions need to be asked. The aim of this initial screening is to identify beforehand situations where it is usually not worthwhile to undertake the efforts to develop new guidance. This may be either because the resulting guidance will only address a marginal problem or will be outdated by the time it is produced or because there are logistic, managerial or organisational barriers that cannot be overcome. The screening questions involve judgement that cannot be entirely objective but that should be made explicit and transparent.

Figure 5. Flow chart for the initial screening for the need of guidance.



Note: the decisions knots in this flowchart have been adapted from the existing WHO handbook[1] and other sources[28].

a. Priority topic

If the topic is not a priority, it certainly does not merit the efforts involved in producing guidance. A 'priority' means a problem that is significant for its magnitude or consequence, and that falls within the mandate of WHO (or eventually, in the mission of any organisation embarking in producing guidance). Ideally, institutions should aim to have a regularly revised plan for priority topics that establishes a menu of problems to be addressed through guidance. In the absence of such a plan, prioritisation will have to occur whenever the opportunity to develop guidance arises.

Some of the criteria to be considered when determining whether the topic is a priority include[29]:

- one or more parameters of the problem to be addressed is remarkable in terms of extension (e.g. number of people affected) or severity (i.e. consequences for the health status of people or the health system);
- the consequences of disregarding the problem may be potentially enormous in a foreseeable future;
- the problem falls into your institution or department's scope of work (i.e. a priority 'for' the institution);
- it may merit special attention when compared with other problems that require consideration.

b. Addressable with guidance

The production of guidance may not be the needed approach when the problem does not entail choosing between different courses of action. For example, if the problem is the establishment of standards, formulating operational procedures may help; or if the issue has to do with the skills of staff undertake certain tasks, training programmes may be more appropriate; or if an action is very accepted and unlikely to do harm. Guidance is needed when:

- there is not a widely accepted way to address the problem yet;
- it is conceivable that the problem can be solved (i.e. tractability, vulnerability);
- the problem cannot be solved by reinforcing existing policies, regulations or procedures[17] (unless the method to reinforce them is part of the problem being addressed);
- there is demand from national policymakers for guidance on the topic;
- there are a number of policy options for addressing the problem

c. Checking whether guidance already exists

If the topic can be addressed with guidance, then a search of the literature and consultations with experts may help to identify existing guidance that could be adapted.

i. Sources of health systems guidance

The table below briefly introduces some sources of guidance, classified in two groups: sites with a focus on health systems and/or guidance and generic literature sources.

Table 5. Sources of guidance.

Source	Institution	Search	Area	Geographical focus
Health systems / guidance sites				
National guideline clearinghouse[30]	US Department of Health and Human Services – Agency for Health Care and Quality	By topic, organisation[*]; advanced search[†]	Clinical, public health	Worldwide
Health Systems Evidence[31]	McMaster University	Policy briefs; by health system topic, Canadian priority areas; type of question; author; dates; updates; open search.	Health systems	Canada
NICE[32]	NHS-National Institute for Health and Clinical Excellence	By type[‡], topic[§], date	Clinical, public health	UK
EVIPNet policy briefs[33]	WHO-EVIPNet	By subject[**], year, language	Any	LMIC
Research policy[34]	WHO	By topic, by year, by title	Any	LMIC
Generic sources				
Health Evidence Network (HEN)[35]	WHO. Regional office for Europe	Type of publication; keywords.	Any	Any
Virtual Health Library (VHL)[36]	(search portal)	By type, clinical aspect, full text, type of study, main subject, people, journal, language, year	Any	Any
Trip database[37]	-	By hierarchy of evidence (includes guidelines by geographical area); 'developing world' switch.	Clinical	Any
NHMRC[38]	National Health and Medical Research Council	By topic	Clinical, research health systems	Australia
NICHSR[39]	US National Library of Medicine	Free, by category, scope	Any	Worldwide

[*] Extensive list of organisations from several countries.

[†] Age of target population, clinical specialty, guideline category, implementation tools, intended users, institute of medicine care need, institute of medicine domain, methods used to analyse the evidence, methods used to assess the quality and strength of the evidence, methods used to formulate recommendations, certain guidelines components, organisation, organisation type, publication year, sex of the population.

[‡] Cancer service guidance, clinical guidelines, diagnostic guidance, interventional processes, medical technologies guidance, public health guidance, technology appraisals.

[§] Following the human body systems, and public health.

[**] Policy Making, Health Programs and Plans, Financing, Health, Child Welfare, Infant Mortality, Acquired Immunodeficiency Syndrome, HIV, Evidence-Based Medicine, Delivery of Health Care, Pre-Eclampsia, Health Policy, Planning and Management, Health Services Administration.

Source	Institution	Search	Area	Geographical focus
Evidence best practice for public health[40]	University of Massachusetts Medical School	Links to other websites	-	-

LMIC: Low- and middle- income countries.

The VHL is a portal that retrieves references from a wide range of data bases, including the WHO library[41], which is especially useful for an initial scoping of the literature. It may be helpful as well to look at some additional sources[42] searched by authors in the Cochrane EPOC group[43]: CINAHL[44], DHSS-Data[45], EMBASE[46], Research and Development Resource Base[47] and Sigle[48].

ii. Search strategy for guidance

Guidance developers may want to outsource the search for existing guidance or engage a collaborator within the group with experience in literature search. Building an effective search strategy that is equally sensitive (not to miss relevant documents) and specific enough (not to retrieve too many irrelevant ones) is quite a specialised job. An informal search with experts who can point at existing guidance may also help. The Cochrane Handbook provides extensive information on how to conduct searches[49].

When building the search strategy a list of essential concepts has to be elaborated. The term 'guidance' will be one of the essential concepts, since this is the type of document we hope to retrieve. The other essential concepts will relate to the problem at hand and possible options of interest. The next step is to unfold the essential concepts with synonyms that are likely to be found in the literature. Synonyms within a single concept are linked with the operator 'or' and groups of essential concepts with the operator 'and'. A simple example is shown below:

Table 6. Example of simple search strategy for guidance on human resources for health in low- and middle-income countries

Guideline or Guidelines or Guidance or Recommendation	and	Developing country or Low- and middle-income country or Poor country	and	Health workforce or Human resources or Staff

The specific syntax of the search strategy will need adaptation depending on the search engines and the sources. MeSH terms should also be considered if appropriate.

The term 'recommendation' is included in the 'guidance' concept because interestingly, in the past when search strategies for guidance were tested, one known and relevant document[20] could not be found using terms with the root 'guid' because it did not contain any term with this root. However, the word 'recommendations' was in the title. Other possible variations or synonyms may need to be considered to avoid missing relevant documents.

The year of publication may be relevant if it is dated, since older guidance is most likely outdated and missing potentially relevant evidence produced more recently. This restricts this search strategy. However, it still could be informative to look at old guidance.

Annex 8 shows an example of sources and hits from a generic search for health systems guidance.

iii. Relevance, inclusion and exclusion

Titles and abstracts available from the search hits will be quickly scrutinised for relevance; i.e. against the terms of the search strategy. For example, some of the terms from the search strategy could be used in different contexts or for different purposes irrelevant to the search of interest (e.g. 'supervision' is used to describe an intervention in the area of human resources management or as 'supervised medical treatment').

Examples of inclusion criteria for documents judged relevant are:

- consist of guidance, considering options to address a problem;
- problems are consistent with the problem as framed by the guidance steering group;
- the type of populations considered in the guidance is consistent with the type of the populations of interest;
- health systems settings are consistent with the types, tiers or parts of the health systems of interest;
- the geographical context or the types of countries are consistent with those of interest;
- guidance is not outdated considering the most recent research on the topic.

iv. Critical appraisal of included guidance

The guidance documents that are included by applying the suggested (or similar) criteria need to be appraised for their quality. The guidance steering group should have decided beforehand how to handle low quality guidance or at which threshold of quality guidance will be considered acceptable. When substandard guidance is identified, further information may be sought from the authors. Some guidance can still be of value if updated or adapted; otherwise it can be discarded straightaway.

The current widely accepted tool used to assess the quality of guidance is AGREE II. Although AGREE II was constructed with clinical guidance in mind, most of the items can be used or adapted for health system guidance. See Chapter 9 for a complete explanation of the AGREE II assessment items.

d. Teams and programmes to develop guidance

If guidance is not available, it should be determined whether other teams in the same organisation, in partner organisations or elsewhere are already preparing guidance relevant to the same topic. If the development of guidance seems justified at this stage, it should be decided where to base the work by ascertaining if programmes that could facilitate the production of guidance exist and if resources and time are available, or can be made available.

e. Availability of resources

A final consideration is the availability of resources: whether they are available or can be made reasonably available. In section 1.3 some considerations about the resources to

develop guidance were anticipated and are fully developed in the chapter dealing with management (Chapter 10).

Note:

The existence of evidence in the form of systematic reviews is considered a screening criterion in the current GRC handbook. However, this means that evidence has to be searched before deciding to develop guidance. In the field of health systems guidance, it will be more difficult to find systematic reviews so guidance developers may have to consider commissioning systematic reviews or basing the recommendations on primary research.

2.2.2 Outputs: decision matrix

The whole screening process needs to be documented. A form can be used to record the events and the final decision. This can be especially useful as a reference document for teams that may explore guidance development for the same topic in the future.

The figure below proposes a form to document the screening process to decide on the development of guidance. It contains all decision items in the screening process and the final decision.

Figure 6. Form for screening for the need for guidance.

SCREENING FOR THE NEED FOR GUIDANCE

Complete the shadowed areas below.

Item	Explanation		Yes	No
Topic	State the topic (even if provisional) for guidance development			
Problem				
Priority	Magnitude, extension, consequences, institution scope...		☐	☐
Brief justification:				
Needs guidance	The topic or problem has to be addressed mainly with guidance, as opposed to manuals, operating procedures...		☐	☐
Brief justification:				
Missing guidance	Guidance of acceptable quality that addresses the topic is not available (Yes = missing guidance)		☐	☐
	Number of guidance documents retrieved			
	Number of guidance documents included			
	Number of guidance documents of enough quality			
No one else is developing guidance	In your institution or organisation, in others... (Yes = you would be the only team developing guidance)		☐	☐
Brief justification:				
Can fit in a programme	A programme that addresses similar topics, that may have activities related to guidance development.		☐	☐
Brief justification:				
Time and resources available	Are they available or is it possible to make them available so as guidance is developed when required?		☐	☐
Brief justification:				

If there is at least one 'no', consider abandoning the idea of developing guidance

If all are 'yes', consider the possibility of developing guidance for the topic

If guidance is available, the guidance steering group may consider updating or adapting existing guidance. If guidance is not available but the need to revise is an urgent priority, new methods should be developed. Hence, the research topics outlined below.

Research topic 2. Methods for guidance adaptation.

WHO lists several issues to be taken into account when adapting guidance[1]:

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

There are methods to appraise guidelines, such as AGREE[50], guidelines standards[51] and procedures to adapt clinical guidelines[52]. However, there does not seem to be a validated approach to adapt health systems guidance to different contexts or for health systems interventions under consideration when they do not fully match interventions researched (e.g. in terms of users, health system setting, relevance or outcomes or contextual issues).

Research topic 3. Rapid methods for guidance development.

The timing when guidance needs are formulated, guidance is developed and decisions are taken is essential to ensure that the best available evidence is used for decision making[53]. The time required for those events and their sequence are essential aspects to facilitate the uptake of guidance when it is required.

Rapid methods imply that not all evidence to support guidance can be accessed, appraised and included in the guidance development process. Still, guidance needs to take into account both evidence on outcomes and on implementation (a single systematic review typically takes more than one year) and to be free of bias. Some examples of rapid guidance exist (e.g. for specific situations or conditions[54]), rapid assessments[55]; and the use existing guidance that can be endorsed by WHO[12].

3 How to scope health systems guidance

What if before starting doing what ought to be done
we start doing what we should have done?
(Quino)

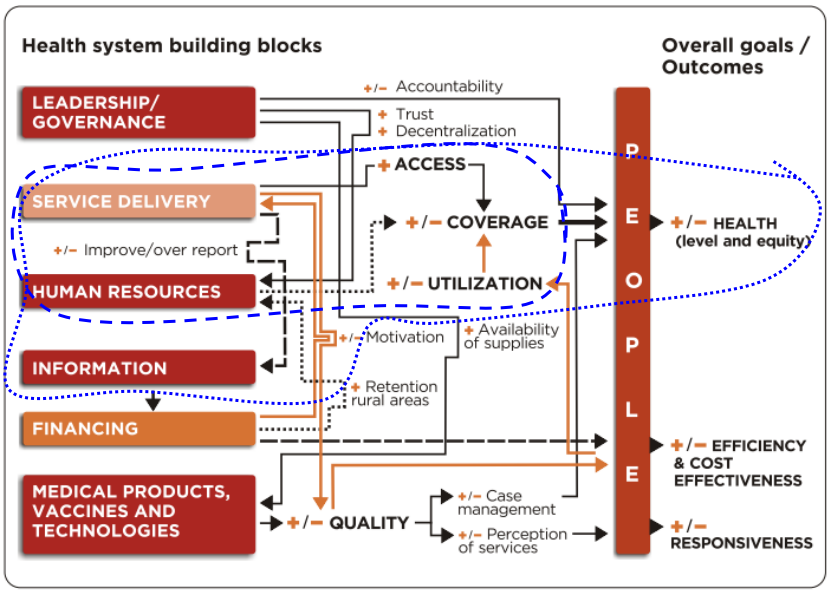
At this stage, the problem is framed and the need for new guidance, based on the problem and other criteria, is justified. The first section addresses the scoping of guidance: the definition of the parameters that will determine how guidance is developed and disseminated (e.g. geographical scope, audience, decision level). The second section deals with the production of the protocol, which will be the basis for the methods followed in the next steps and will serve to obtain institutional clearance for developing guidance.

3.1 Approach: scoping guidance

'Scoping' entails setting boundaries to the potentially unlimited spectrum of issues that could be covered by guidance on a specific topic. The scope of guidance will be the basis for the next steps in the guidance development process (i.e. searching for evidence).

The scoping of guidance should follow the parameters which defined the problem. The model used to frame the problem should also support the scoping of guidance. In the example of Figure 7, two slightly different scopes for guidance are proposed based on the same model (two interrupted blue lines). In one case the building block of 'information' and health status and equity outcomes are taken into account, in the other they are not.

Figure 7. Conceptual pathway for the 'pay for performance' intervention with two possible scopes.



If the scope for guidance is too general or inclusive it may lead to an unnecessary workload and to conflicting or inconsistent evidence that is difficult to articulate into sound recommendations. On the other hand, if the scope is too narrow it is more likely that no evidence is found, especially in health systems topics where research synthesis is not so profuse.

3.1.1 PICOT question(s)

Once the problem has been framed it needs to be 'translated' into questions. In other words, the generic question of 'how can problem X be solved' has to be shaped in a way that could be used to

search for evidence. While systematic reviews are articulated around 'interventions' (i.e. what are the effects of intervention X?), guidance is articulated around the problem (outcomes) (i.e. what interventions can solve problem X?). The problem to be addressed by guidance is clear at this stage of guidance development as it has already been framed. However, interventions of interest have not yet been established.

The acronym PICOT ('Participants – Interventions – Control – Outcomes – Time') is used to frame effectiveness questions in the context of primary research and research synthesis. It is reasonable to use it for scoping guidance as well. First, the problem as framed in the preceding chapter will be expressed in terms of outcomes. Then, interventions that may have an impact on those outcomes will be selected.

Since guidance is built from the problem framed, several or even many interventions dealing with the problem will have to be scrutinised. In other words, in the development of guidance often more than one or even many PICOT questions will have to be formulated and addressed.

a. Participants (P)

These are the subjects who are affected by the problem and/or the main factors associated with the problem. Following the major types of outcomes (population and health system); it would be useful to also consider population and health systems participants:

(i) Population participants

They represent the sector of the users interacting with the health system who are affected by the problem or by its ultimate consequences (e.g. rural population). However, not everyone is equally susceptible or affected by a given problem: equity is key for any sound health policy, as health inequities arise from social conditions and systems' incapacity to deal with illness[56]. Several approaches have been suggested to address 'equity' in the context of primary research and research synthesis[57]):

- class, status, education, occupation, income and assets, gender, race, ethnicity, caste, tribes, religion, national origins, age, and residence (Measurement and Evidence Knowledge Network);
- place of residence, race/ethnicity, occupation, gender, religion, education, socioeconomic status, and social capital, plus other factors (Campbell and Cochrane Equity Methods Group and the Cochrane Public Health Review Group; acronym PROGRESS plus).

(ii) Health system participants

Participants within the health system: those who are at the centre of the health system problem, such as staff, certain types of health facilities, or a combination of both. Health system participants can be defined on the basis of the health systems framework and the problem modelled.

The guidance steering group will decide to which extent the types of participants need to be defined when searching for evidence. At one extreme, any participant may be considered. Later on evidence can then be analysed by sub-groups of interest from the evidence available in the systematic reviews retrieved, if possible. At the other extreme, only specific groups of participants are included in the search for evidence at the expense of reducing the chance to retrieve relevant systematic reviews. This requires careful judgement based on the problem that actually needs to be addressed as well as on other features, such as geographical scope or decision level of guidance.

In any case, participants have to be clearly specified and different types of health systems arrangements (e.g. public, private) should be kept in mind. For example, "participants could include users and non-users of health services, as well as institutions such as health facilities, where utilisation data could have been collected. Study designs could also have used facilities or districts as units of allocation. We included studies on all types of providers (governmental or private). We did not limit the scope of our study to a particular level of healthcare delivery and all types of health services could be included in this review"[58].

b. Interventions and comparators (I, C)

As mentioned above, the starting point for health systems guidance is the problem for which guidance is needed (i.e. guidance on what to do, on what interventions address the problem). Once the problem has been modelled and the ways to measure it defined as outcomes, the guidance steering group will have to figure out which interventions would be worthwhile to consider. Some interventions may be obvious but others are not, especially if interventions are innovative, 'imported' from other domains or only recently tested.

Interventions have to be properly defined, as it is done in systematic reviews: the formulation of interventions should state what they consist of, what are the key elements (e.g. as opposed to other similar interventions). For example: "user fees, e.g. the introduction or removal of fees or variation in the level of fees. User fees must be understood here as a financing mechanism that has two main characteristics: payment is made at the point of use of the health service, and there is no risk sharing. User fees can entail any combination of drug costs, supply and medical material costs, entrance fees or consultation fees. They can be paid for each visit to a healthcare provider or encompass a whole episode of illness"[58].

There does not seem to be a systematic classification of interventions matching specific outcomes. However, there are several ways to devise interventions that could be relevant to the framed problem:

- use the modelled problem to identify the components of the health systems related to the problem, and then to identify interventions centred on those parts of the systems;
- look at existing taxonomies of interventions and consider to which extent interventions listed may have an impact in the problem of interest. There are several taxonomies on health systems interventions available (see Annex 5)[59,60,61]; for example:
    - professional
    - consumer
    - organizational
    - financial
    - regulatory

  or:
    - governance arrangements
    - financial arrangements
    - delivery arrangements
- informal scoping of the literature;
- consultations with experts.

Interventions of interest should be consistent with the problem framework; i.e. it should be made explicit how interventions could reasonably have an impact on the outcomes and, eventually, on the problem addressed by guidance.

Comparators may include current practices or alternative interventions as described in the literature. It may not always be possible to define comparators, specially when the topics under consideration can hardly be studies using experimental designs.

c. Outcomes (O)

Outcomes are 'measurable' aspects of a problem that may refer to health status or to health system indicators. They will be used to retrieve evidence, to justify recommendations and to suggest evaluation approaches for the policy options outlined in the recommendations. Outcomes should be identified and drawn from the modelled problem.

(i) Population level outcomes

The most relevant outcomes for the population are those related to their health status. Even in the case of interventions which are not strictly clinical, it has been argued that interventions need to also show effects on clinical outcomes rather than only on intermediate outcomes (e.g. interventions to enhance medication adherence[62]). However, it is not always possible to relate an intervention to measurable changes in health status outcomes, especially in health system-level interventions.

If it does not seem feasible to consider health status outcomes, other types of population level outcomes can be considered, such as household costs for care, health seeking behaviour practices or satisfaction with health care. This should not be seen as a deterrent to issue recommendations. It is reasonable to assume that the cumulative improvements in the health system would have an impact on the population health status, even if this cannot be directly measured.

(ii) Health system level outcomes

Health system level outcomes refer to measurable changes in specific parts of the system and should be identified in the framework used to model the problem. There does not seem to be a unique approach to look at possible health system outcomes.

The scheme 'input-processes-outputs[23]' is widely used in the areas of quality of care[63*] and monitoring and evaluation[64], and could be helpful in selecting health system level outcomes. One way to proceed could be to combine the components of the health system (e.g. the WHO health systems building blocks) with the 'inputs-processes-outputs' scheme. Following this approach, Table 7 provides some examples of possible 'outcomes' of interest.

[*] 'Outputs' can be understood as long term health status outcomes or as the immediate results from activities. The relations between changes in inputs, processes and outputs are not necessarily simple and direct.

Table 7. Examples of health systems inputs, processes and outputs indicators.

Building block and example of event		Inputs	Processes	Outputs
Leadership / governance	Decentralisation process	Implementation of SWAPP	Multi-sectorial approach	Consistency between decentralisation of competencies and budget
Service delivery	Outreach vaccination	Mobile cold chain	Multi-year vaccination plan	Outreach vaccination sessions
Human resources	Lay health workers	Resources for training, candidates	Training of LHW	Availability of LHW
Information	Strengthening of HMIS	Availability of software	Standardisation of data management	Timeliness of data outputs
Financing	Health insurance	Initial capital	Administrative procedures	Insured people
Medical products	Stock management scheme	Software, staff	Training, roll out of scheme	Stock-outs

LHW: Lay Health Worker. SWAPP: Sector Wide Approach.

An analytical model that puts population and health system outcomes in a structured relation[65] establishes a succession of events from the availability of services up to effectiveness:

- Availability
- Accessibility
- Acceptability
- Contact
- Effectiveness

In order to define specific outcomes, these events can be articulated around plausible causal relations with the main problem of interest.

It is anticipated that the evidence to support health system recommendations may be relatively weak in most cases and, therefore, that recommendations will have to be carefully pondered in relation to contextual and implementation issues. Thus, it would be sensible to accompany recommendations with suggestions on how to monitor and evaluate their implementation. This would provide policy makers with a backstopping mechanism, allowing them to make decisions early interrupting or scaling up a course of action (see Chapter 7). Even at this early stage, it would be sensible to consider outcomes from the perspective of what may be worthwhile and feasible to monitor and evaluate and not only as the effects of interventions, which are defined under the research paradigm.

The checklist below[64] may be useful to assess the appropriateness of outcomes, from the perspective of what may need to be evaluated:

- The outcomes and their indicators are specific, measurable, achievable, realistic and time-bound (SMART).
- The outcomes clearly outline an area of work where the agency and its partners can have significant influence.
- The outcomes are worded in such a way that they communicate what has changed, for whom and by when.
- The outcomes clearly address the interests and concerns of men, women and marginalized groups.
- The outcomes address changes in institutional capacities and behaviour that should lead to sustainable development of the country or region.
- The outcomes speak to changes in conditions and capacities and not delivery of products and services.
- The outcomes have indicators that signal how the desired change will be measured.
- The outcome indicators are measures of change that go beyond what one agency will produce or deliver.
- The outcome and its indicators provide a very clear and precise image or picture of what the future should look like, and is not so general that it could cover almost anything.

The existing WHO handbook proposes creating a list of outcomes and rating their importance by the members of the guidance steering group as a method to choose the outcomes to be considered for recommendation. In the preceding paragraphs we suggested to link 'outcomes' to problems, to consider population and health system outcomes, and then to identify those outcomes that are also relevant in the evaluation of recommendations. A scoring system could also be useful in the final stages of deciding which outcomes to choose.

d. Time (T)

The definition of the time frame in which effects on outcome should be observed is specially relevant in health systems guidance, due to the relevance of sustainability issues. It is usually easier to demonstrate effect on the short run that may easily vanish after some time.

3.1.2 Primary and secondary audiences (decision level)

Guidance has to do with recommendations on courses of action that may involve any tier of the health system: from central Ministries of Health up to remote rural health posts; but the primary audiences of guidance are those who are involved in the policy making process. The site for policy decisions is the 'higher' level of the system, either governments or international agencies. Governments can issue policies at the national level and, in decentralized settings, at the sub-national level, where a 'government' with policy competencies exists. Decisions at lower levels of the health system should comply with national policies and be more strategic or operational in the lower tiers of the system.

All those involved in the policy decision making process and in implementation are considered to be the primary audience of guidance: policy makers, groups of users and other stakeholders. The primary audience is involved in the guidance development process as members of the guidance development group.

The secondary audience is comprised of the remaining constituencies who may have some interest in the problem. This includes researchers (e.g. to identify research gaps), groups with interests in similar problems or approaches, or stakeholders working in wider health systems and development issues. Secondary audiences have an interest in guidance but do not use it for decision making.

The needs of different constituencies are addressed (1) when setting up the processes to issue recommendations (Chapter 6) and (2) when planning for the dissemination of guidance (Chapter 7.2.3).

3.1.3 Additional issues

The guidance steering group may want to restrict the scoping of guidance using additional criteria. For example, the geographical scope: defined as a certain homogeneous socio-cultural region (e.g. guidance limited to the Latin-American and Caribbean region), by a certain national income level (e.g. low-income countries), or by any other criteria.

In the scoping of guidance, it will be helpful to provide clear definitions of the terms used. Many terms relevant to the area of health systems have lay and technical meanings that can be easily confounded (e.g. 'system', 'inequality', 'inequity', 'results based financing').

3.2 Outputs: protocol and initial approval

The guidance scope establishes what falls within and beyond the work of the guidance steering group and the other supporting groups and experts. The full protocol for guidance development is articulated around the scope.

The protocol details the approaches and methods that will be followed in developing guidance. In the method section, it should detail:

◊ the search strategy to retrieve evidence that will be used to inform guidance development;
◊ the sources searched;
◊ inclusion and exclusion criteria
◊ the quality criteria to assess the quality of systematic reviews;
◊ and the data extraction template.

The methods that comply with the terms of the scope are defined beforehand because guidance has to be unbiased (i.e. consistent with the evidence available) and transparent (i.e. everybody can follow what has been done, and how). Post hoc arrangements are discouraged. For example, adapting the outcomes considered relevant to the outcomes for which evidence is available. Elaborating a protocol before evidence that supports guidance is retrieved and made available reduces the impact of authors' biases, promotes transparency of methods and processes, and allows peer review of the planned methods[66].

Table 8 suggests the structure of a protocol for developing health systems guidance.

The protocol is compliant with the WHO's GRC procedures. It is proposed that it is included in the documentation to be submitted to the GRC for initial approval of the guidelines development. The following items are the documents to be submitted to the GRC for initial approval of the guidance development process:

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

All items in the ‘planning clearance form’ are part of the protocol, although the protocol also details the methodological aspects to be followed.

Table 8. Table of contents of the protocol to develop guidance and links to the handbook chapters.

	Protocol sections	Brief description	Chapter or section in handbook
1.	Guidance dashboard	Single summary page with the items listed below.	-
	a. Topic and problem (summary)	Clear statement of the problem.	2.1
	b. Guidance scope (summary)	Guidance scope: PICOT, audience and additional issues.	3.1
	c. Guidance steering group	Members and clearance of conflict of interest.	10
2.	Problem framed	Problem and model used to frame it.	2.1
3.	Rationale for developing guidance (need for guidance)	Justification of why new guidance is needed.	2.2
4.	Methods:		
	a. Search methods for identification of reviews	Sources of evidence and search strategies; timeframe of searches and justification.	4.1.1
	b. Inclusion / exclusion criteria for selecting evidence	Criteria for including evidence based on the scope of guidance with explicit definitions consistent with the scope of guidance.	4.1.2
	c. Methods to assess the quality of systematic reviews	Quality of reviews and how to proceed depending on the quality of reviews.	4.1.3
	d. Data extraction	How data will be extracted and missing data handled (e.g. missing data on outcomes of interest).	4.2
	e. Methods to assess the quality of evidence	Quality of evidence on outcomes and interventions and presenting evidence for issuing recommendations.	5
	f. Approach to elaborate recommendations	Rationale and methodology to develop recommendations.	6
	g. M&E approaches for recommendations	Suggested approaches to evaluate recommendations, where appropriate.	7
	h. Guidance dissemination plan	Detail of which means will be used to make guidance available to the primary and secondary audience.	7.2.3
	i. Methods to appraise the quality of guidance and follow up	Appraisal of guidance produced and how to elaborate and follow up or guidance update plan.	9
5.	Management (including declaration of interests and internal and external sources of support)	All issues that have to do with the management of the processes to develop guidance in the institution.	10
6.	Glossary	Key concepts used all along the protocol, especially in the scope of guidance.	Annex
7.	Acknowledgements	-	-
8.	References	-	-
9.	Protocol review history	Details on the review(s) of the protocols and changes made.	-

Note: adapted from the "Cochrane Handbook for Systematic Reviews of Interventions"[49]. M&E: Monitoring and Evaluation.

4 How to search and retrieve evidence for guidance on health systems

L'expérimentateur qui ne sait point ce qu'il cherche
ne comprend pas ce qu'il trouve
[The researcher who does not know what he is searching for
does not understand what he finds]
(Claude Bernard. Rapport sur les progrès et la marche de la physiologie générale en France. Paris, Imprimerie impériale 1867)

If development of guidance on a given topic is justified (section 2.2), the scope of guidance has been defined (3.1) and the methods detailed (section 3.2), the next step is to search and retrieve the research evidence that will support guidance*.

Research evidence is retrieved to support recommendations ("a summary of all relevant research evidence is essential when developing recommendations"[1]). As will be further discussed in Chapter 6, recommendations are the result of combining evidence on both the effects of interventions (for the outcomes of interest) and on the feasibility of implementing those interventions.

Consequently, the evidence needed to produce guidance includes:

a. evidence on the effects of interventions on outcomes:
   - evidence on intervention(s) and co-intervention(s) addressing the problem;
   - evidence on effects modifiers and sub-groups of participants;

b. evidence on implementation issues (i.e. operations issues).

What constitutes evidence on the effects of interventions and on implementation issues is a matter of fine judgment. Implementation refers to the adoption of research findings into practice[67]; it has to do with the practicalities of how interventions can be set up, while taking into account resources, timing, organisations, management and any other factor that may affect their deployment. All implementation issues have to be taken into account in any case of guidance development, although not all issues will be equally relevant in all cases.

In contrast, co-interventions refer to those additional or complementary changes in the arrangements of the health system or its components that are required for the main or critical intervention(s) to take place. While implementation issues can be seen as routine, most of them are of a managerial nature. Co-interventions imply structural changes, the adoption of new strategies or the abandonment of existing ones.

In the clinical field this distinction can be more straightforward because co-interventions can be seen as health system-level interventions needed to deliver new diagnostic tools or treatments. For example, women affected with breast cancer could be offered breast-conserving surgery plus radiotherapy or mastectomy, when appropriate. Since women from disadvantaged groups are less likely to be offered a choice, evidence on interventions to overcome barriers is retrieved and combined with the evidence about the clinical intervention[68]. Evidence on barriers would be searched by looking at co-interventions that are necessary to delivery the main intervention to a certain population group. Evidence on barriers is likely to be related to health system-level interventions or interventions at the interface between the health system and the population.

In health systems guidance, 'main intervention(s)', co-interventions and implementation issues have to do with the health system. The 'modelling of the problem' (see section 2.1.1) should assist the guidance steering group in deciding what the main interventions, co-interventions and (unspecific)

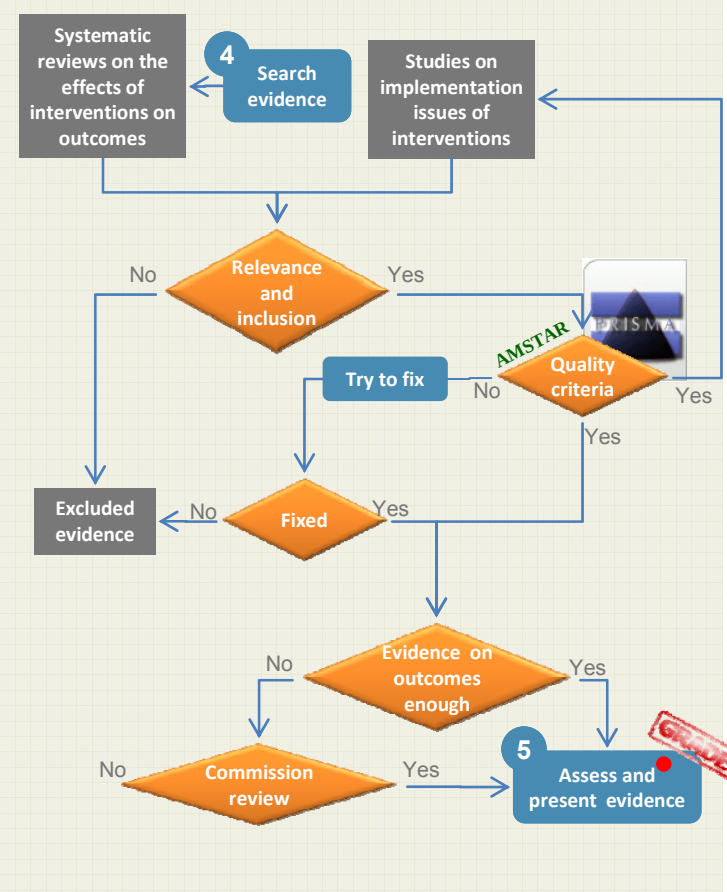
* Colloquial evidence has also its place in guidance development but, by definition, it is not retrieved using specific systematic methods.

implementation issues are. For example, a human resources intervention ('main intervention') may need a new financial arrangement within the health sector ('co-intervention'); and implementation issues would have to do with the organisation capacity to deploy those interventions and costs considerations.

Evidence on both the effects of interventions and on implementation issues has to be retrieved. The best sources of evidence are systematic reviews, which are the product of research synthesis*, and are widely used by guidance developers[69]. Research synthesis is a systematic, transparent and reproducible method to synthesise all available research evidence on a given topic[70], minimising bias by considering the quality of the underlying studies.

Figure 8 describes the steps involved in gathering research evidence to support guidance: beginning with the search and retrieval of systematic reviews and ending with their presentation to the group of panellists issuing recommendations. Both evidence on the effects of interventions and on implementation issues are represented in the figure.

Figure 8. Events related to the search, inclusion and quality assessment of systematic reviews.



* Although systematic reviews –and meta-analyses- can be used to provide evidence on a wide range of issues (e.g. defining a problem), in this step of guidance development only systematic reviews on the effects of interventions and on implementation issues are relevant.

4.1 Approach: searching, including and assessing evidence

4.1.1 Searching for evidence

Searching the literature is a complex task: there are many potential sources of evidence, each one with specific requirements for being accessed and queried. Searching and retrieving studies is a specialised activity that is much better carried out by a professional. The guidance steering group should seriously consider collaborating with a search specialist to retrieve the evidence needed for guidance development.

Two search strategies should be considered: one to retrieve evidence on the effects of interventions on outcomes and another one to retrieve evidence on implementation issues, although they do not take place simultaneously (see Figure 8).

a. Evidence on the effects of interventions

Following the WHO handbook for guidelines development[1], it is suggested to use existing evidence from systematic reviews synthesis on the effects of interventions as a source of evidence, where possible.

Systematic reviews on the effects of interventions are structured around problems, participants, interventions and outcomes. They are not framed as policy questions, which are about needs, services and programmes or strategies. Therefore, a complete match between the policy question and the research evidence is hardly possible. There are several potential problems in using reviews[71] that need to be taken into account:

Table 9. Inconveniences and way around in relation to the use of systematic reviews to support guidance.

Inconvenience	Way around
The ways reviews are framed around research topics can hardly match the guidance scope.	Searches for evidence can refer to several interventions to improve matching to the guidance scope.
When data is pooled, it may include participants, interventions or outcomes which could be irrelevant and excluded when developing guidance.	Sub-group analyses*[72]; access to the individual studies included in the reviews.
Often lack relevant information on implementation issues.	Implementation issues are explicitly considered in the guidance development process.

The search for reviews is based on the scoping of guidance. The basic syntax of the search strategy consists of a series of terms linked with the operator ‘and’, and each term having a series of synonymous or variants linked with the operator ‘or’ (see Table 10).

* Sub-group analyses have to be undertaken with caution. See reference for further details.

Table 10. Example of generic and simple search strategy for guidance on human resources in low- and middle-income countries

Study		Participants		Intervention(s)		Outcomes		Other
Systematic review	and	Health care providers or Traditional healers or Infants and children	and	Lay health workers or Staff incentives	and	Adherence or Access or Re-admission	and	Neonatal care or Child health

The scope of guidance will dictate to which extent each part of the search strategy needs to be developed. For example, guidance may be about any intervention to address the problem of interest, in which case there would be no terms limiting the search to a specific intervention or groups of interventions. On the other hand guidance may be restricted to specific areas of health care (e.g. child health or vaccination).

One of the essential terms refers to the type of study of interest; in this case, 'systematic reviews'. Evidence on the effects of interventions is better found in systematic reviews of experimental studies which precisely aim at assessing to which extent an intervention produces effects on certain outcomes. Therefore, the primary source of evidence will be systematic reviews of experimental or quasi-experimental studies. The Cochrane Effective Practices and Organisation of Care group that deals with health system level interventions includes controlled before and after and interrupted time series studies, besides randomised controlled trials, in their systematic reviews[73]. Reviews (or even single studies) including other study designs can be considered as well, especially if the topic is unlikely to be addressed using experimental study designs (e.g. financing mechanisms), or when there is compelling evidence from multiple, large, well designed observational studies.

Other terms in the search strategy follow the PICOT acronym used to scope guidance (see Section 3.1). The interventions to be considered are those that have been designed to address the problem (as defined in the scope of guidance). Interventions which may not be critical to the problem but that may have an influence on how other interventions work (i.e. 'co-interventions') should also be included. For example, interventions around training are often considered alongside other main interventions targeting human resources, such as supervision.

The scope of guidance may have some additional items, such as a focus on particular population groups, equity or ethical considerations. When systematic reviews contain such information, no additional search is needed. However if, as is often the case[97], systematic reviews are not responsive to those issues, additional searches will need to take place.

The guidance steering group should elaborate a structured list of terms following the logic of Table 10 and based on the scoping of guidance. While the logic of the search strategy is unique, the specific terms and terminology to be used has to be adapted to the specific requirements of each database or portal searched. The generic syntax is submitted to the literature search specialist to be translated into the specific syntaxes for each literature source to be searched. The guidance steering group may consider piloting alternative search strategies to balance sensitivity (i.e. many hits with higher likelihood of finding irrelevant ones) and specificity (i.e. less hits with higher likelihood of missing relevant ones).

b. Evidence on implementation issues

The distinction between transferability and applicability can be useful to address implementation issues[74]. Transferability describes whether the effectiveness of the programme would be similar to the level detected in the research settings; and applicability whether the intervention process could be implemented in the local setting. Several frameworks exist to identify relevant implementation issues, in terms of implementation[75] and applicability[76].

Evidence on implementation issues can be retrieved form several sources:

- systematic reviews assessing the effects of interventions may describe implementation factors related to the interventions being tested;
- systematic reviews of other study designs (e.g. qualitative research);
- primary research.

Although there are useful frameworks to sort out implementation and context issues that should be included from the stage of conducting systematic reviews[77], evidence on these issues is not as available in systematic reviews[78]. Systematic reviews focused on implementation issues are also scanty because these are highly contextual issues, making synthesis and generalisation more difficult. It has even been argued that systematic reviews may be unhelpful to address issues that pertain to local evidence (e.g. local evidence about on-the-ground realities and constraints, the values and beliefs of citizens, interest group power dynamics, institutional constraints, and donor funding flows)[98].

The ultimate goal of implementation research is the production of evidence which would allow some generalization beyond particular situations, enabling predictions about how well interventions would work given particular conditions[79*]. This evidence is often meagre because systematic reviews on the effects of interventions tend to obviate this type of evidence (partially because it may be missing from the primary studies included in the review) or because evidence on implementation issues is better addressed by studies that require other types of research synthesis, which are not so well developed[80] (e.g. synthesis of qualitative studies). However, there are actually several methods and examples to synthesise this type of evidence[80].

It is advisable that the plethora of implementation issues that can be relevant for each intervention be delimitated to ensure that the search for evidence is achievable in the time frame of guidance development, without overlooking any relevant aspects. The search for evidence on implementation issues should be guided by:

- the scope of guidance;
- the implementation issues identified in reviews on the effects of interventions (as suggested in Figure 8, evidence on implementation issues is searched after evidence on the effects of interventions on outcomes has been retrieved);
- a predefined list of implementation issues, as the one in Table 24 (section 5.2), summarised below:
    - Intervention as tested: costs, appropriateness, ease, testable.
    - Facilitators: credibility, skills, experience, size of team, resources, stability.

[*] Several different aspects are included in the term ‘implementation research’: research on ways to uptake research findings by policy makers and putting them into practice, research on how to apply findings from basic sciences into the clinical field, and research into organisational, systemic and contextual issues.

- Users' organisation: priority, capacity, timing.
- Context: political, socio-economic, cultural, administrative, other sectors, donors, other stakeholders, people's rights.

Implementation issues can also been framed using a series of five questions[75]:

1. what are the barriers to implementation?
2. what are the strategies in planning the implementation of a new policy in order to facilitate behavioural changes of users?
3. … and behavioural changes of providers?
4. … and organisational changes?
5. … and system changes?

The syntax of the search strategy parallels the one used for systematic reviews on the effects of interventions (Table 10). Study design terms can be left open, because implementation issues can be reported in many different study types, such as in observational studies, qualitative studies and even reviews. The guidance steering group may consider including qualitative research in different ways[81]:

- to discuss and interpret quantitative findings, to help make sense of it, or to place the review findings into context;
- to undertake a review of qualitative studies using formal qualitative synthesis to interpret the findings of the quantitative synthesis (parallel synthesis). Researchers might choose to include qualitative research embedded within the quantitative studies or stand-alone qualitative studies that address the question of interest;
- to combine the results of quantitative and qualitative syntheses. This approach is sometimes referred to as multi-level, sequenced, cross-design or meta-synthesis.

The search strategy should be adapted to keep the size of this task manageable and careful judgement must be applied to decide what evidence is reasonable enough. Limiting terms could include (a) the specific interventions under consideration and (b) a few terms related to the key implementation issues.

Table 11. Example of generic, simple search strategy for implementation issues.

Study		Intervention(s)		Issues		Limiters		
Systematic review or Stakeholders analyses or Meta-ethnography	and	Lay health workers or Staff incentives or Training	and	Costs or Equity or Stakeholders	and	Implementation or Scaling up	and	Low- and middle-income countries or Scaling up

c. Equity

Equity considerations are relevant for both the evidence on the effects of interventions and the evidence on implementation issues. The term 'equity effectiveness' designates the efficacy being modified by access, coverage, diagnostic accuracy, and provider and patient adherence[82] and, eventually, other factors. The 'evidence-based framework for equity-oriented knowledge translation'[83] expands on the implications of considering equity in knowledge translation initiatives (i.e. for guidance development). According to this framework, values, awareness, and resources for the 6 Ps (public (community), patient,

press, practitioner, policy-maker and private sector) have to be assessed by socio-economic status and modifiable barriers by socio-economic status have to be prioritised.

The guidance steering group may want to consider querying systematic reviews on the effects of interventions by applying an equity lens (see section 4.2). However, if equity related terms are included in the search strategy, it is likely that some potentially relevant reviews can be missed, either because equity terms are not used for indexing the reviews or because they have not been addressed at all. Even in the latter case, it can be useful to retrieve those reviews to access the primary research that may contain explicit equity considerations.

d. Evidence on costs

Costs can be considered as outcomes of the interventions of interest and as an implementation issue. 'Costs' is used as a generic term to point at the financial and economic implications of interventions and are an intrinsic part of guidance and guidance development[1,84].

Economic evaluations constitute in themselves a whole body of knowledge and methods that can easily overcome the efforts gone into health systems guidance development itself. Evidence on costs can be found in systematic reviews, alongside findings on the effects of interventions, in single studies of economical evaluations or in modelling studies. There are many types of evaluations that could be relevant in the context of health systems guidance development: incremental cost-effectiveness ratios, cost-utility analyses, cost-effectiveness analyses, cost minimisation analyses, cost-benefit analyses and cost-consequence analyses, among others[85]. Costs considerations are key in ensuring that recommendations are sound, adaptable to specific contexts and feasible to implement. However, searching, selecting and appraising evidence on costs is a specialised task that warrants the inclusion of an economist as part of the guidance steering group[84].

The extension and depth of costs considerations will depend on the scope of guidance and to which extent costs have been identified as a critical issue in the interventions under consideration. The judgment of the guidance steering group will define how far or how deep to go.

In practical terms, the minimum requirement for WHO guidance is to link resource use to the interventions under consideration[1,86]. Evidence on costs should include qualitative and quantitative information on the following costing parameters, where appropriate:

- cost of providing the intervention(s) and to handle adverse events;
- costs of implementing the intervention(s), setting them up and scaling them up;
- costs incurred by the people, the health systems and the society;
- costs in terms of finances and also time and opportunity costs;
- costs with a time perspective and for all interventions (or comparators) being considered.

e. Sources and hits

There are different types of sources of research evidence that the guidance steering group can consider:

- Repositories of systematic reviews; e.g. Cochrane and Campbell collaborations.
- Literature databases; e.g. Medline.
- Experts on the topic for guidance.
- References in existing guidance related to the topic (see section 2.2.1).
- Websites

Annex 9 lists a good number of sources, including specific thematic sources (e.g. to retrieve economic evaluations[87]). The scope of guidance (i.e. topic, problem, health system components) will determine which sources are worthwhile to explore. An experienced search expert can very well assist in the selection of sites. The guidance steering group should also decide on the timeframe in which to search for each source. There can be grounds for limiting the timeframe of searches although this has to be properly argued and documented in the protocol.

Sources and search strategies have to be explicitly reported in the protocol as well as in the documentation of guidance development. Hits from several sources are usually stored as text files, either in delimited or fixed width formats. Hits should be stored as retrieved and saved, to keep track of the whole process and as safeguard in case mistakes are made in the future processing of references. For example. the outcomes of the searches should be documented with a table containing all sources searched, a link to the search strategy used for each source, the dates for which references have been searched, the date when the search was done, the number of hits and the name of the file where hits were stored (see Table 12).

Text files should be loaded into 'reference manager' software which would allow one to:

- store all hits into a single database;
- structure the information according to standard citation fields (i.e. author, title, year, …);
- distinguish between different sources;
- identify and mark duplicate references;
- safely import and export references list;
- mark additional information (e.g. included, excluded…).

There are several commercial reference manager software such as 'Reference Manager'[88], 'Endnote'[89] or 'Procite'[90]. All of them are available for purchase, although there is a free Endnote web version[91]. For a comparison of software see F. Dell'Orso[92].

Other software (e.g. electronic spreadsheets, databases management software) could also be acceptable. However, the functions needed may require some manipulation or even programming (e.g. macros, filters).

Table 12. Literature search hits and documentation.

Sources	Search strategy	Date 1	Date 2	Date searched	Hits	File
Pubmed	Search A1	1965	04/2011	01/04/2011	508	A1_Pubmed.txt
VHL Cochrane	Search A2	1990	04/2011	01/04/2011	301	A2_Cochrane.txt
…		…	…	…	…	…
VHL WHOLIS		1990	04/2011	01/04/2011	(117)	A2_WHOLIS_1.txt
VHL WHOLIS		1990	06/2011	01/06/2011	214	A2_WHOLIS_2.txt
Total hits for Searches A (effects of interventions):						
Pubmed	Search B1	2000	04/2011	01/04/2011	415	B1_Pubmed.txt
…	…	…	…	…	…	…
Total hits for Searches B (implementation issues):						

Notes. Dummy figures. Searches A: search strategy to retrieve evidence on the effects of interventions; Searches B: search strategy to retrieve evidence on implementation issues. Figure in brackets is not counted in the total number of hits because a later search of the same database was conducted in this example.

Once all references are stored into a single database, duplicates will be identified and marked as such (for the sake of completeness, no records should be deleted). Care should be taken to keep track of potentially relevant hits from which to extract evidence on the effects of interventions and on implementation issues. For example, one study may be unrelated to the former because it does not contain evidence on what does or does not work, but it may address an implementation issue (e.g. costs) applicable to the interventions being considered.

Special circumstances should be explicitly and fully documented, such as language or publication status restrictions or other circumstances that influence the number of hits retrieved.

4.1.2 Relevance and inclusion of evidence

A first screening for relevance of documents retrieved can be done by looking at their titles and abstracts. The objective of this initial screening is to identify obviously mistaken hits, such as documents that are not even related to the topic of guidance. This can be done by one person only and should be over-inclusive, keeping rather than discarding doubtful records. Then, the database containing the references should be updated accordingly, marking records that are deemed irrelevant.

Those documents which are relevant are then retrieved in full text and formally assessed for inclusion, according to the scope of guidance. An explicit file naming and storage system should be designed for electronic and hard copy document maintenance. The filing system should contain a numbering method consistent with that of the documents in the reference database.

An inclusion / exclusion form should be prepared following the items in the scope of guidance; namely:

- type of document (e.g. systematic review, overview; or other types of studies to gather evidence on implementation issues);
- interventions of interest;
- participants, components of the health system;
- outcomes of interest, if defined;
- implementation issues, if defined;

Each criterion will be assessed as 'done' (or 'met', or 'yes'), 'not done' (or 'unmet' or 'no') or 'unclear' or 'doubtful'.

Two separate inclusion / exclusion processes are carried out since criteria for both types of evidence are different: (1) for evidence on the effects of interventions and (2) for evidence on implementation issues. Documents will be classified as 'included' if all criteria are fulfilled; as 'doubtful' or 'unclear' if none of the criteria is unmet, as 'unclear' if at least one of the criteria is met; and 'excluded' if none of the criteria is met. Ideally, two members of the guidance steering group should independently assess the inclusion / exclusion of documents. Disagreements would be sorted out as predefined in the protocol, usually by a third member of the guidance steering group and/or by discussion.

In order to properly apply these criteria, clear definitions should be attached to them and written down in the formats used to assess inclusion / exclusion. These definitions should be consistent with the ones given in the scoping of guidance.

Table 13. Table to assess and record inclusion and exclusion criteria of documents (partial snapshot).

INCLUSION / EXCLUSION

	Study					Inclusion criteria				
	Id number	Author	Year	Status	Agree	Study	Participants	Intervention	Control	Outcomes
	Effects of interventions on outcomes									
1	0092	Abc	2010	To assess	X					
2	0197	Def	2000	Include	OK					
3	0198	Ghi	2001	Include	X					
4	0245	Jkl	2004	Exclude	OK					
5	0319			To assess	X					
6	0392			To assess	X					
	Implementation issues									
1	0738			To assess	X					
2	0745			To assess	X					
3	0797			To assess	X					
4	0838			To assess	X					
5	1730			To assess	X					
6	3470			To assess	X					
7	4240			To assess	X					
8	0018			To assess	X					
9	0052			To assess	X					

Several formats can be used, both electronically and in hard copies, to support the process of inclusion / exclusion. A 'list' in an electronic support (e.g. a spreadsheet) has several advantages: it can be automatically produced from the references database, it eases browsing along all relevant documents, it allows automatic inclusion / exclusion decisions based on the two independent

assessments (i.e. using formulas) and it is relatively compact in terms of file and storage space (see an example in Table 13).

The inclusion / exclusion process also documents the reasons for exclusion, which should be reported in the final guidance, both for evidence on the effects of interventions and on implementation issues.

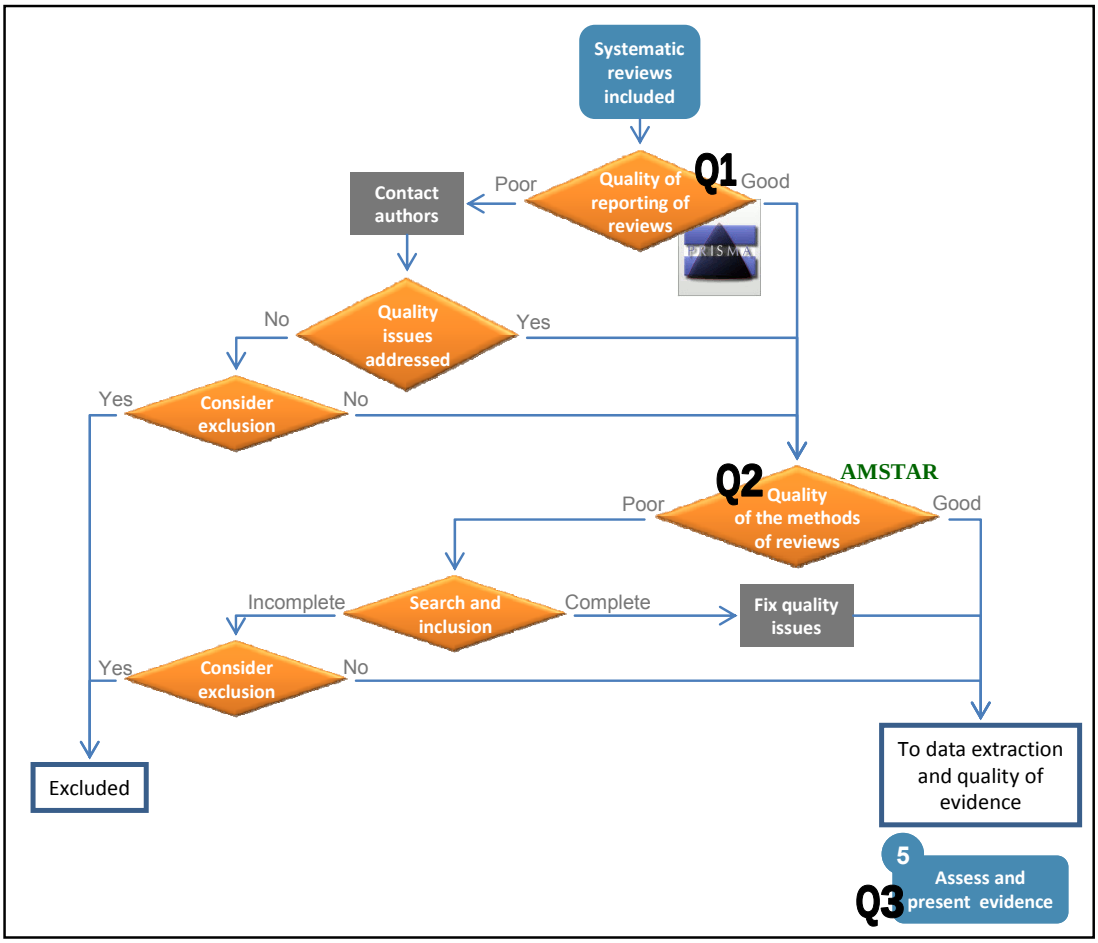
4.1.3 Quality of included documents

Included documents have to be assessed for their quality. The guidance steering group will have decided in advance how to proceed with poor quality documents: whether to exclude them or to consider them 'second class' evidence. This should have been explicitly stated in the protocol. The problem of including poor quality reviews is that evidence from these reviews (and the recommendations drawn from that evidence) can be biased (and the degree of bias cannot be estimated and it can be potentially very large).

There are three types of quality to be considered when assessing the evidence reported in systematic reviews: the reporting quality of the systematic review (arbitrarily called 'Q1' in Figure 9), the methodological quality of the review ('Q2') and the quality of the evidence ('Q3') in the reviews (the latter will also depend on the reporting quality of the underlying primary research studies and on the quality of the evidence in the primary research studies, see section 5.1). To some extent, the methodological quality cannot be properly assessed if the reporting of the methods is not exhaustive.

Figure 9 represents the flow of decisions and actions when assessing the quality of the included reviews. If the quality of reporting is not adequate to assess the quality of methods, the authors of the reviews can be contacted for clarifications. If the quality of the reporting issues cannot be properly addressed, then the guidance steering group may consider excluding the review. Otherwise, the methodological quality of the review is assessed. Again, there may be methodological quality issues. However, if the search and inclusion strategies of studies in the review is comprehensive, then the guidance steering group may consider addressing any remaining quality concern themselves, thereby avoiding redoing the review from scratch. However, if the quality concerns relate to the search, retrieval and inclusion of the studies processes, trying to amend this would require a similar effort to conducting a review from the start.

Figure 9. Flow chart related to the quality of reviews.



A widely accepted tool for assessing the reporting of systematic reviews is the PRISMA[93] statement (Transparent Reporting of Systematic Reviews and Meta Analysis). The AMSTAR[94] (Assessment of Multiple Systematic Reviews) tool is used to assess the methodological quality of systematic reviews. Some items overlap, as shown in Table 14. The guidance steering group may first consider undertaking an assessment of the quality of reporting and second an assessment of the methodological quality of the review. However, in order to simplify the process, one could directly measure the methodological quality of the review and then elucidate if any existing methodological concern is due to a problem in reporting or true methodological pitfall. There are other tools to asses the quality of reporting (the Institute of Medicine Standards for Reporting Systematic Reviews[95], Annex 10) and the methodological quality (see CEBM in see Annex 10) of systematic reviews. The Equator network[96] contains a repository of reporting guidelines which include many types of study designs*.

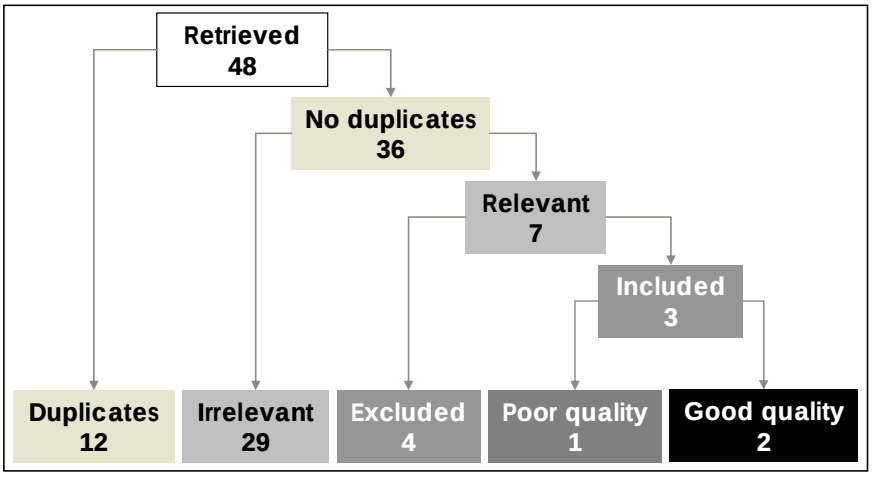
Quality assessments are undertaken independently by two assessors, as was done when deciding on the inclusion or exclusion of documents. Quality assessments have to be documented. A spreadsheet can be used, in which all included documents are listed. The columns are used for the quality items; either for the quality of reporting, for the methodological quality of reviews or for both,

* Experimental studies, observational studies, diagnostic accuracy studies, reliability and agreement studies, systematic reviews, qualitative research, economic evaluations, quality improvement studies, reporting data, sections of research reports, specific conditions or procedures.

as chosen by the guidance steering group. The quality assessment can be recorded in the same types of sheets used to retrieve evidence from documents in the form of quantitative and qualitative data (see section 4.2).

The final classification of the documents retrieved is also reported in the form of a tree indicating how many of them have been discarded or included in each step of the process. An example is provided in Figure 10.

Figure 10. Tree reporting the number and status of systematic reviews retrieved.



Note: dummy data.

Table 14. Comparison of PRISMA and AMSTAR tools.

Item	PRISMA		AMSTAR
Title	1	Identify the report type (systematic review, meta-analysis, both)	(nothing)
Summary	2	As applicable: background; objectives; data sources; study eligibility criteria; participants and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications; registration number	(nothing)
Rationale	3	Rationale for review in context of what is already known	(nothing)
Objectives	4	Addressed questions with reference to participants, interventions, comparisons, outcomes, study design (PICOT)	(nothing)
Methods			
Protocol, registration	5	Availability, accessibility and full registration of review protocol	1	'A priori' design including research question and inclusion criteria*
Eligibility criteria	6	Study and report characteristics used to assess eligibility, with rationale	4	No exclusion based on publication type, publication status or language
Information sources	7	All information sources; date of last search	3	More than one electronic database searched. List of databases and years; search complemented by consulting current contents, reviews, textbooks, registers, experts, references in identified studies†
Search strategy	8	Full electronic search strategy including limitations for at least one database	3	Key words and/or MeSH terms; search strategy if feasible
Study selection	9	Process for study selection	2	Two independent study assessors, consensus procedure for disagreements
Data collection	10	Data extraction and confirmation / supplementation methods	2	Two independent data extractors, consensus procedure for disagreements
Data items	11	List and definition of all variables; assumptions and simplifications		(nothing)
Risk of bias in individual studies	12	Methods to identify bias and use of bias assessment outcome in data analysis	7	Scientific quality assessed with 'a priori' or other applicable methods

* Some iteration may be needed overall in reviews that are not about effectiveness.

† The date of the searches can be especially relevant to ensure that evidence is not out of date and that the most recent studies have been included in the reviews.

Item		PRISMA		AMSTAR
Summary measures	13	Statement of principal summary measures		(nothing)
Synthesis of results	14	Methods of data handling and summary, including measures of consistency for each meta-analysis	9	Ensure studies are combinable, assess homogeneity*. Appropriateness of combination and combination method
Risk of bias across studies	15	Methods used to identify bias	10	Graphical and statistical evaluation of publication bias
Additional analyses	16	Describe additional analysis methods, indicate if pre-specified		(nothing)
Results				
Study selection	17	Number of studies screened, assessed, included. Justify exclusions. Provide flow diagram	5	List of included and excluded studies
Study characteristics	18	All studies: characteristics for all extracted data. Provide citation	6	Provide data on participants, interventions and outcomes, and range of characteristics of included studies
Risk of bias within studies	19	Risk of bias assessment outcome; outcome-level assessment if available	7	Document scientific quality of included studies
Results of individual studies	20	For all outcomes and studies: summary data for each intervention group and effect estimates with confidence intervals and forest plot		(nothing)
Synthesis of results	21	Results of each meta-analysis, including confidence intervals and consistency measures		(nothing)
Risk of bias across studies	22	Risk of bias assessment outcome	10	Outcome of graphical and statistical evaluation of publication bias
Additional analysis	23	Report results of additional analysis		(nothing)
Discussion				
Summary of evidence	24	Summarize main findings including strength of evidence for each outcome; consider relevance		(nothing)
Limitations	25	Limitations at level of studies, outcome and review	8	Consider limitations in study methodological rigor and scientific quality in analysis, conclusions and recommendations.
Conclusions	26	Interpretation in context of other findings; research implications		(nothing)
Funding	27	Funding source, additional support, role of funders for systematic review described	11	Source of support acknowledged in systematic review and included studies

* For non effective reviews, ‘combination’ may be purposely because different types of studies contribute to different understandings of the topic.

4.2 Outputs: extraction of information

The documents included and assessed constitute the whole body of research evidence that supports guidance. Evidence from those documents (reviews and may be other types of studies) has to be extracted in a systematic way to inform recommendations.

4.2.1 Information on the effects of health systems interventions

One set of information is the synthesis of the effects of interventions across all reviews. This can be quantitative and/or qualitative. Quantitative data can be presented as pooled estimates (e.g. as a result of a meta-analyses) or as a series of indicators from different studies. Qualitative evidence is typically presented in a narrative way and may point at plausible scenarios in which interventions may induce changes on certain outcomes, even in the absence of robust quantitative evidence.

The information of the effects of interventions will feed into the 'outcomes profiles' (see section 5.1) and will constitute one of the 'axes' for issuing recommendations (the other axis being evidence on implementation issues).

This information is best retrieved using a grid to enter data extracted from the reviews. Since the interest in guidance development is not only to show what does or does not work, but also to provide explanations and options, it may be advisable to extract data on the individual studies apart from the pooled estimates of the effects of interventions.

Typically the types of information to extract are:

a. Reference to the document containing the data.
b. Outcomes description.
c. Technical specifications of the measurements (e.g. what methods were used to measure the outcomes).
d. Statistics (e.g. type of statistic parameter used, values), only for quantitative data.
e. Values (the specific measurements in figures or in narrative form).
f. Interpretation and limitations.

These types of information are further detailed in the table below:

Table 15. Generic and implementation issues information to be extracted from included documents.

Heading		Information items	Heading		Information items
Reference	1	Family name of first author of the review (if pooled data)	**Statistics**	11	Parameter (e.g. percentage, odds ratio, relative risk)
	2	Year of publication of review (if pooled data)		12	Adjusted (e.g. for clustering)
	3	Family name of first author of the study	**Values (for intervention**	13	Number of subjects in each group
Outcome description	4	Code	**and for controls)**	14	Point estimate
	5	Text of the outcome		15	Confidence intervals
	6	Type of outcome (e.g. health status, behaviour)	**Interpretation and limitations**	16	p values
Measurement	7	Subjects (e.g. certain types of patients, providers)		17	Narrative (for qualitative data)
	8	Measuring instrument or tool		18	By the authors
	9	Time of measurement in relation to intervention start		19	By the guidance steering group
	10	Participants age range			

This information is used to design a data extraction sheet where all information related to the effects of interventions on outcomes will be entered (Table 16). Ideally, data should be independently entered by two members of the guidance steering group and inconsistencies between both data entries checked and resolved.

Care has to be taken not to double count evidence from any single study that may have been reported in two different publications or that can be included in more than one systematic review, since this evidence would count double and artificially increase its 'weight' among the other findings.

Table 16. Example of data extraction sheet for the effects of interventions (snap shot of a partial view).

Stu			Author and year								Group 1						Group 2					
	Out	Stu/Out	Text of outcome	Category	Subjects	Measurement	Time	Stat1	Stat2	Adjust	Stat1	Stat2	CI	Low	Upp	N	Stat1	Stat2	CI	Low	Upp	N
001			**Hallstrom 2004**							**Yes**	**CPR Only**					**107**	**CPR plus AED**					**128**
	01	001/01	Survivors of definite cardiac arrest	Health status	Residential units	Clinical		N			1.0					37	1.0					33
	01	001/01		Health status	Public units	Clinical		N			14.0					70	29.0					95
	02	001/02	Survivors of definite or uncertain cardiac arrest	Health status	All	Clinical		N			16.0						31.0					
	03	001/03	Cerebral performance	Health status	Survivors of definite arrest	Clinical		N	%		10.0	71.4					22.0	73.3				
	03	001/03						N	%		3.0	21.4					5.0	16.7				
	03	001/03						N	%		1.0	7.1					3.0	10.0				
002			**Pilote 1996**								**Usual care**					**79**	**Monetary incentive**					**82**
	01	002/01	Completion of 1st follow-up appointment	Behaviour	All	Records		N	%	No	42.0	53.0	95.0	47.0	59.0	79	69.0	84.0	95.0	76.0	92.0	82
	01	002/01																				
005			**Sullivan-Bolyai 2004**							**No**	**Intervention**					**22**	**Control**					**20**
	01	005/01	Parental concern	Social	All	Score	0 mont	Mean	SD		159.0	31.0					136.0	37.0				
	01	005/01			All	Score	1 mont	Mean	SD		139.0	25.0					142.0	39.0				
	01	005/01			All	Score	6 mont	Mean	SD		136.0	26.0					130.0	42.0				
	02	005/02	Maternal confidence	Social	All	Score	0 mont	Mean	SD		37.0	6.0					38.0	5.0				
	02	005/02			All	Score	1 mont	Mean	SD		38.0	5.0					38.0	5.0				
	02	005/02			All	Score	6 mont	Mean	SD		41.0	4.0					41.0	6.0				
	03	005/03	Impact on family	Social	All	Score	0 mont	Mean	SD		63.0	10.0					41.0	6.0				
	03	005/03			All	Score	1 mont	Mean	SD		59.0	11.0					56.0	11.0				
	03	005/03			All	Score	6 mont	Mean	SD		58.0	13.0					59.0	13.0				
	04	005/04	Home care resources	Social	All	Score	0 mont	Mean	SD		31.0	5.0					33.0	5.0				
	04	005/04			All	Score	1 mont	Mean	SD		33.0	5.0					32.0	6.0				
	04	005/04			All	Score	6 mont	Mean	SD		34.0	5.0					33.0	5.0				

Adapted from the work done for the lay health workers review[108].

4.2.2 Generic and implementation issues information

The other axis (besides the evidence on the effects of interventions) to support recommendations consists of information on implementation issues. Generic information on implementation issues is used to elaborate the interventions profiles (see section 5.2).

The types of information to extract include:

a. Document characteristics: citation information (main author, year of publication, title, source), author affiliation.
b. Objective and topic: objectives of the study, topic areas (e.g. using existing health systems frameworks).
c. Methodological characteristics: study design; if a systematic review: technical characteristics of the search (databases, grey literature, dates…), geographic scope, level of the health system, specific limitations.
d. Methodological quality, with the items assessed, depending on the quality assessment tool.
e. Participants: description of those who are the beneficiaries of the intervention(s).
f. Providers: those who actually deliver the intervention(s).
g. Interventions (and comparison, if any): describing the interventions involved and the comparators; also co-interventions.
h. Outcomes: types of outcomes, synthesis of evidence in quantitative (e.g. meta-analyses) or qualitative (e.g. narrative meta-ethnographic synthesis) forms.
i. Implementation issues: costs, individual and organisational capacity, contextual issues.

Table 17 shows an example of a list of items used to retrieve generic and implementation issues information on health systems level interventions. A comprehensive list showing the information items to be extracted should be part of the protocol for guidance development and be the base for the data extraction sheets. Each item of information should include a clear definition, hints of where to find the information in the full text documents, how information has to be entered (e.g. codified or as literal quotations from the text) and space for additional comments to assist the data retrieval process.

Table 17. Generic and implementation issues information to be extracted from included documents.

Heading		Information items	Heading		Information items
Document characteristics	1	Family name of first author of the review	**Participants**	37	Service provider
	2	Year of publication of review	**Providers**	38	Service provider Other
	3	Main author' organisation		39	Qualification providers
	4	Country of main author's organisation		40	Qualification of providers, other
	5	Country of main author's organisation in code		41	Participants clients
	6	Country region		42	Participants other
	7	Country income group		43	Participants age range
Objective / topic	8	Topic detail	**Interventions Comparison**	44	Intervention, detail
	9	Topic		45	Intervention
	10	Primary objective		46	Delivery strategies
	11	Service delivery HS block		47	Delivery strategies, detail
	12	Workforce HS block		48	Control
	13	HIS HS block	**Implementation**	49	Costs
	14	Products HS block		50	Individual capacity
	15	Financing HS block		51	Organisational capacity
	16	Leadership HS block		52	Co-interventions
	17	Other topics or themes		53	Stakeholders support
Methodological characteristics	18	Design	**Context**	54	Name / type of program, project...
	19	Number of studies included		55	Qualification providers
	20	Type of included studies		56	Qualification of providers, other
	21	Databases searched		57	Support received
	22	Grey literature searched		58	Delivery strategies
	23	Search eldest year		59	Delivery strategies, other
	24	Search most recent year		60	Socio-economic context
	25	Geographical scope		61	Political context
	26	Level of the system		62	Other
	27	Language			
	28	Other criteria			
	29	Limitations			
Methodological quality	30	Outcome, detail			
	31	Outcome			
	32	Adverse events			
	33	Stat methods and tests			
	34	Potential sources of bias, potential confounders			
	35	Quality of includes studies assessed?			
	36	Methodological quality other			

Adapted from X Bosch-Capblanch, “The GAVI Knowledge Bank” project. HS: health system.

From an equity perspective[97], it could be useful to specifically look for the following items in systematic reviews*:

1. a logical model suggesting ways that the intervention can work and how equity factors may influence its effectiveness;
2. definitions of 'disadvantage' and for whom the interventions are intended, considering the presentation of findings by sub-groups;
3. study designs that are consistent with their "fitness for purpose" (not merely following an evidence hierarchy);
4. outcomes reported for specific categories of 'participants' (e.g. PROGRESS plus[102]), as defined in the scoping of guidance (see section 3.1):
    - **P**lace of residence
    - **R**eligion
    - **O**ccupation
    - **G**ender
    - **R**ace/ethnicity
    - **E**ducation
    - **S**ocioeconomic status
    - **S**ocial networks and capital
    - Plus: socio-economic status, age, disability, sexual orientation, other vulnerable and socially excluded groups.
5. process evaluation information on the factors influencing the effectiveness of the intervention in certain groups;
6. outcomes in disadvantaged populations;
7. a discussion of the applicability of findings in real-life situations.

Since reviews may not have extracted all relevant information on implementation issues from the included primary research, it is advisable to extract information from individual studies included in the reviews as well.

* These items have been recommended to be included in systematic reviews in the first place. However, it is likely that a number of systematic reviews do not explicitly address most of them.

Table 18. Example of data extraction sheet for generic and implementation issues information (partial snap shot partial view).

Review	Study		Topic		Countries	WHO HS Bulding blocks								
Author Year	Author Year	Type	Topic	Topic detail	Countries	Geographical scope	Rural / urban setting	Service delivery HS block	Workforce HS block	HIS HS block	Products HS block	Financing HS block	Leadership HS block	Other topics or themes
Ejemot 2008		Review	Diarrhoea	Hand washing for preventing diarrhoea	#N/A	NA	#N/A	Unclear	No	No	No	No	No	NA
Ejemot 2008	Bartlett 1988	Study	Diarrhoea	Diarrheal illness among infants and toddlers in	USA	Part of a district	NA	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Black 1981	Study	Diarrhoea	Handwashing to prevent diarrhea in day-care	USA	Part of a district	Urban	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Bowen 2007	Study	Diarrhoea	effect of a handwashing-promotion	CHN	Part of a region	Rural and urban	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Butz 1990	Study	Diarrhoea	Occurrence of infectious symptoms in	USA	Part of a district	Urban	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Carabin 1999	Study	Diarrhoea	reducing infections in toddlers	CAN	Part of a region	NA	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Haggerty 1994	Study	Diarrhoea	Community-based hygiene education to	ZAR	Part of a region	Rural	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Han 1989	Study	Diarrhoea	Prevention of diarrhoea and dysentery by	MMR	Part of a district	Urban	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Huang 2007	Study	Diarrhoea	handwashing in the prevention of diarrhoeal	USA	NA	NA	Yes	No	No	No	No	No	health behaviour
Ejemot 2008	Kotch 1994	Study	Diarrhoea	hygienic intervention in child day-care	USA	Part of a district	NA	Unclear	No	No	No	No	No	health behaviour
Ejemot 2008	Ladegaard 1999	Study	Diarrhoea	Hand-hygiene and sickness among small	DNK	Town, city	NA	Unclear	No	No	No	No	No	health behaviour

Adapted from the GAVI knowledge bank project.

4.2.3 If there is not enough evidence

The guidance development process is based on a written protocol describing the approaches and methods used in each step. One of the utilities of the protocol is that, at this stage, missing relevant evidence can be easily identified. In the protocol, relevant information items were detailed in terms of participants, interventions, outcomes and key implementation or contextual issues to be considered. It should now be clear that some of the information that was considered relevant for developing guidance may be missing.

What can be done?

- Explain the judgement used to determine whether evidence is sufficient, and to distinguish between lack of evidence and evidence of no effect.
- Evidence gaps have to be clearly reported. It is as important to describe evidence on a certain issue as it is to state that for a certain issue there was not enough evidence. This can be recorded in the final guidance document (see Table 37).
- Consider to which extent it is likely that evidence actually exists but was not retrieved. If it is likely that evidence exists, then the search strategy can be modified. This has to be done systematically and transparently as per protocol. Therefore, the protocol should be modified and this modification should be clearly stated and considered for approval. The search strategy could be adapted by (a) broadening the terms (e.g. more synonymous); (b) including additional sources; (c) considering more study types.
- Evidence on the effects of interventions addressing similar problems may be used as indirect evidence for the current guidance.
- Data available for individual studies may allow for additional analyses, such as sub-groups analyses. However, this has to be done with expert statistical advice since it can be a source of error[72].
- The guidance steering group may consider commissioning a systematic review (in Figure 8 following Figure 1). This will entail additional time and resources. Should the group consider basing guidance on a series of individual studies then standard methods should be used, which would be equivalent to conducting a new review anyway[98].
- Recommendations could also be based on what is available even if this is largely colloquial evidence in the form of expert opinions. Recommendations will have to be accompanied by strong monitoring and evaluation considerations to closely observe the effects of the recommendations in real life situations (see 7.1).
- Finally, possibly undertaking new primary research can also be considered (see section 7.2.3).

5 How to assess and present evidence

Everybody knows the dice are loaded
Everybody rolls with their fingers crossed […]
Everybody knows the captain lied
Everybody got this broken feeling
(Leonard Cohen. Everybody knows)

When the former step is finalised, there is no more retrieval of evidence, unless a gap is identified in the following steps. In order to use evidence to inform recommendations, evidence will have to be synthesised and then presented to the team responsible for issuing the recommendations.

Decisions on whether to implement a given intervention depend on two main issues:

1. to what extent the intervention works;
2. to what extent the intervention can be implemented*.

This is the reason why evidence has to be synthesised for those two aspects. The outputs of this step will be, respectively:

1. An outcomes profile on the effects of the interventions (Section 5.1), with comments.
2. An interventions profile on the implementation issues of the interventions (Section 5.2), with comments.

These outputs will be presented to the guidance development team for deliberations and will inform the issuing of the recommendations. The interpretation and use of information depends on the contents as well as on the format. There are several initiatives that present evidence in different formats and for different purposes[99]. These initiatives make evidence accessible and provide a wide variety of ways to present it. Most of them are focused on a single systematic review or present topics for wider dissemination. It is also possible that evidence is already packed in formats such as policy briefs. The challenge to present evidence is to make it usable by the team issuing recommendations: to make the underlying concepts comprehensible, and to facilitate a correct interpretation on its meaning and implications. Hence, for 'outcomes' and for 'implementation' issues, there will be a sub-section proposing a presentation of the evidence. The presentation of evidence has to:

- be centred on the problem of interest, and not necessarily on a single intervention;
- include evidence on outcomes and evidence on implementation issues;
- be unbiased, understandable and user friendly;
- encourage transparency and discussion within the team issuing recommendations.

To assess means "to evaluate (a person or thing); to estimate (the quality, value, or extent of), to gauge or judge"[100]. The assessment is done by identifying the key features of 'outcomes' and 'interventions' and by appraising each one. For example, the observed effects are features of 'outcomes and the costs are features of 'interventions'.

Table 19 and Table 20 provide a list of features that qualify 'outcomes' and 'interventions'; i.e. what has to be taken into account when appraising the effects of interventions (Table 19) and the implementation issues (Table 20). These features have been extracted from the GRADE tool and from ExpandNet items used for planning the scaling up of interventions.

* Including health system setting issues, contextual issues and the characteristics of the policy formulation processes. See also the beginning of Chapter 4.

Table 19. Qualifiers of the effects of interventions with the GRADE items.

Feature		Described with:	Threatened by:	Items in GRADE
DIRECTION AND SIZE OF THE EFFECT				
Is it beneficial or harmful?[101]	**Direction (a)**	• Direction of the effect • p value	• Effect modifiers • Confounding • Publication bias	Baseline risk; Relative and absolute effects; Magnitude; Confounding; Publication bias
How large is it?	**Size (b)**	• Relative and absolute changes[76] and duration		
INTERNAL VALIDTY OF THE EFFECT				
Is it close to the 'true' effect?	**Accuracy (d)**	(Risk of bias has to be appraised)	• Bias	Risk of bias
Is it always the same?	**Precision (c)**	• Confidence interval	• Random error	Imprecision; Precision
	Consistency	• Heterogeneity	• Different populations, interventions, outcomes across studies	Inconsistency
Is the 'cause-effect' true?	**Causality**	• Dose-response gradient	• (same factors above)	Dose-response gradient
EXTERNAL VALIDTY OF THE EFFECT				
Is it relevant?	**Directness[76] (e)**	(Has to be appraised)	• Comparability of participants, settings and context	Indirectness
Is it important?	**Fit to the problem[101,103] (e)**	(Has to be judged)	• Misjudgement	Importance
PARTICIPANTS				
Can the intervention(s) be delivered equitably?	**Equity[57, 97,102]**	• **P**lace of residence • **R**eligion • **O**ccupation • **G**ender • **R**ace, ethnicity • **E**ducation • **S**ocio-economic status • **S**ocial networks and capital • Plus: socio-economic status, age, disability, sexual orientation, other vulnerable and socially excluded groups.	1. Lack of logical model 2. Grouping effects 3. Inadequate study design 4. Wrong outcomes 5. Lack of implementation considerations	

GRADE starts with an initial quality of evidence based on the study design.

Letters in parenthesis show the items that are represented in the graphic of the evidence profile on the effects (Figure 11): (a) position of the bubbles above or below the middle line; (b) distance of the bubbles from the middle line; (c) size of the vertical bars in the bubbles; (d) size of the bubbles themselves; (e) position of the bubbles in the X horizontal axis.

Other lists are available in the literature as shown by the numbers in the table pointing at the references at the end of the document.

Table 20. Qualifiers of the implementation issues of interventions with the GRADE items.

	Feature	Described with:	Threatened by:	Items in GRADE
ISSUES RELATED TO THE INTERVENTION AS TESTED				
Is it costly?	**Costs**[101]	• Costs of implementing • Effects on costs	• Evidence on costs can be biased as evidence on outcomes	Costs Cost (resource utilization)
Is it appropriate?	**Values and preferences**	• Evidence on values and preferences	• Evidence on costs can be biased as evidence on outcomes	Values and preferences
Is it easy to implement?	**Feasibility**	• Change from existing practices • Technical requirements • Organizational arrangements • Resources	• Complexity • Technical sophistication • Conflict in user organizations • Resources needed	
Can implementation be tested?	**Evaluation**	(not stated)	(not stated)	
FACILITATORS OF IMPLEMENTATION (POLICY MAKERS, MANAGERS)				
Is authority credible?		• Profile of the team leader	• Lack of key figure	
Has the team the skills needed?		• Skills to support training, management, service provision, advocacy, human rights and gender perspectives, resource mobilisation, research, monitoring and evaluation	• Lack of skills • Lack of resources to support the team • Too few members	
Has the team the experience?		• Experience in implementation, advocacy and policy development	• Lack of team members with experience • Lack of examples to mirror	
Is the team large enough?		• Size of team in relation to the tasks and workload	• Lack of prevision in the selection of the team	
Has the team the resources to work?		• Resources in relation to the tasks and workload	• Lack of initiative to mobilise funding	
Are members of the team likely to continue over the time frame of scaling up?		• Likelihood of continuity of team members	• Attrition	
TARGETED / USER ORGANISATION (HEALTH SYSTEM ISSUES)[76]				
Is the intervention compatible with current priorities?		• Perceived need • Policy priority	• Lack of testing capacity	
Is there capacity to implement it?		• Perceived need • Leadership, managerial, technical, human resources, infrastructures, equipment[*21] • Capacity test • Impact on other programmes	• Weak capacity • Lack of resources optimisation • Legal and policy framework • Unforeseen capacity constraints • Lack of vision on how to address constraints	
What is the timing and circumstances that makes implementation possible?		• Opportunities and constraints of organisation changes	• Changes in the implementing organisation	
CONTEXTUAL (ENVIRONMENTAL) ISSUES[76]				
Political, socio-economic, cultural, administrative contexts Other sectors Donors Other stakeholders[101,103] Peoples' rights		• Support / opposition in each contextual components • Recruitment of champions and neutralisation of opponents • Related initiatives to expand implementation[76] • Formal and informal connections • Stability of opportunities and constraints while implementing • Monitoring of contextual components	• Lack of institutional support • Lack of individuals support • Lack of flexibility[101,103] in the implementation strategy • Misjudgements in the assessment of the contextual issues.	

* The WHO Health Systems building blocks Framework could be used here.

The features in the tables should contain considerations and are classified as follows:

- Outcomes
    - Direction and size of the effect
    - Internal validity
    - External validity
    - Participants (equity)
- Interventions
    - Issues related to the intervention itself
    - Facilitators of implementation
    - Target or user organisation
    - Context

Table 20 is not necessarily complete. Other issues may include: policy legacies, collective values, donor pressure, issues pertaining to non-health sectors, as well a humanitarian and ethical issues. Further detail and explanations around the table are found in the next sub-sections and in the corresponding annexes.

5.1 Evidence on the effects of health systems interventions

5.1.1 Approach: appraisal of the quality of evidence

Evidence on the effects of interventions in a given outcome has to be appraised to elucidate to which extent it is beneficial or harmful (direction and size), if it is really what it seems to be (internal validity) and if can be of any use in a specific setting (external validity). Direction, size, internal and external validities and participants (equity) are considered when assessing the 'quality of evidence'

These features are generally assessed in systematic reviews. It might be that some reviews do not contain any assessment or only an assessment of some of the features described in Table 20. In any case, the guidance steering group may want to critically review existing assessments in systematic reviews.

The approach to assess the quality of evidence has to be systematic in order to:

- avoid biases in the assessment of the risk of biases;
- make the assessment reproducible by other teams, or later on in the course of updates of evidence or guidance;
- make it transparent so different people can discuss it;
- identify gaps in the features of evidence which may have been poorly described or not described at all.

Currently, the most widely accepted tool to appraise the quality of evidence of effectiveness is GRADE. The GRADE approach[103] has all the attributes to make it systematic. This approach separates the quality of evidence from the strength of recommendations by recognising that recommendations have to be drawn from both the 'effects' of the interventions and from the features of the interventions that make them relatively feasible to implement. The 'units of analysis' of GRADE are the outcomes: assessments are done outcome by outcome, rather than for all outcomes in a given intervention.

GRADE assessments, which can be done using their free software available online[104,105], have several advantages over using non-specific software. For one, it is readily formatted to host all relevant information needed for the quality assessment (e.g. statements of the research question, type of study, quantitative data). Additionally, it has a tree structure which allows grouping of individual outcomes under profiles, calculations although simple are readily done, and it can

produce different types of outputs (e.g. GRADE profile, summary of findings table). Lastly, assessments can be exported to formats compatible with the Cochrane Library software (Review Manager), although the guidance steering group may not need this latter feature.

The output of the GRADE assessment of the quality of evidence is as follows[106]:

Table 21. Definitions of the GRADE levels of the quality of evidence of effectiveness.

Grade	Definition
High	We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

The grading is based on a starting level of quality which depends on the type of study reporting on a given outcome: randomised trials ('high') or observational studies ('low'). Then, the grade is increased or decreased depending on the assessment of the quality features. There are five features that can lower the quality of evidence (in deep orange colour in Table 19 and Table 22): limitations in design (risk of bias), inconsistency, indirectness, imprecision and publication bias. Conversely, there are three features that can increase quality (in green colour in Table 19 and Table 22): large effect, the fact that the plausible confounding would reduce the demonstrated effect or increase the effect if no effect was observed and a dose-response gradient. (The blue colour in those tables refers to GRADE criteria for the strength of recommendations and will be addressed in another Section).

Table 22. GRADE criteria for the quality of evidence and strength of recommendations

GRADE
1. Quality of evidence (see below)
2. Strength of recommendations
2.1 Quality of evidence
2.1.0 Study design
2.1.1 Reducers of the quality of evidence
• Limitations of the study design or execution (risk of bias) • Inconsistency of results • Indirectness of evidence • Imprecision • Publication bias
2.1.2 Increasers of the quality of evidence
• Large magnitude of effect • All plausible confounding would reduce the demonstrated effect or increase the effect if no effect was observed • Dose-response gradient
2.2 Balance between desirable and undesirable effects
• Importance of outcomes • Baseline risk of outcomes • Relative and absolute effect of an intervention • Precision of the estimates of the effects • Cost
2.3 • Values and preferences
2.4 • Cost (resource utilisation)

See Annex 11 for further details on the GRADE criteria.

As can be deduced from the items in the tables, the GRADE approach offers a very useful output on the quality of evidence but it is relatively complex to conduct. For example, the risk of bias (one of the criteria to downgrade the quality of evidence) entails assessing the allocation concealment, blinding, loss to follow up, selective outcome reporting and other limitations. Applying GRADE software to figures extracted from papers erroneously is like applying an inappropriate statistical test: both lead to nonsensical results. Since the GRADE software can always produce a score if items are filled in, doing the assessment without the proper expertise may give the false impression that the assessment is correct, simply because outputs were produced.

Several issues have to be taken into account when interpreting the quality of evidence:

1) since health systems research is more often done with designs than randomised controlled trials, it is likely that the quality of evidence will start with a 'low' grade and will likely end up as 'moderate' at best:
    - in practice, most of the final grading of the quality of research evidence will fall in the 'low' range; however, within this low range, it could make sense to discriminate between different levels of quality (e.g. low-high and low-low);

- while the risk of bias of non-experimental studies is greater, moderate or low quality does not mean that better quality can be achieved, as it may often happen in the case of health systems level interventions. For certain health systems level interventions, randomised controlled trials would be of less value because many such interventions are system wide or affect such large parts of the system that randomisation is impossible (it may sometimes be theoretically possible to randomise but not practically possible for political reasons);
- if evidence is of low quality, it does not necessarily imply a recommendation to undertake further research, if this research will not (cannot) improve the quality of the evidence;
- finally, a grading the quality of evidence as 'low' may suggest, by contrast, that even lower quality colloquial evidence (e.g. evidence from personal experience, tacit knowledge) which is not based on research, is more valuable, which may not be the case;

2) the technical terminology and statements used are complex even for experienced professionals and involve considerable statistical expertise (e.g. "the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect includes no effect and the upper confidence limit includes an effect that, if it were real, would represent a benefit that would outweigh the downsides");
3) the lay terms, although understandable by all (e.g. 'high quality') actually have a technical meaning, since they synthesise a series of criteria. On these grounds, evidence of 'low quality' may be dismissed, despite its possible value if the underlying scores are positive;
4) the scoring system cannot be interpreted as a magnitude but as an order; e.g. 2 is not double of 1, but it is better than 1;
5) GRADE, by design, cannot assess the quality of outcomes which have not been tested but that can be important when issuing recommendations.

Should GRADE be used to assess the quality of evidence of effectiveness related to health systems levels interventions? Definitively yes, unless there is another more widely and tested systematic approach (see above) to assess the quality of evidence. Experience in developing more systematic health systems guidance will probably unveil limitations of the GRADE approach and propose adaptations to be built on the existing tool and on the expertise gathered so far[107]. There are arguments in favour and against the use of a common approach in grading evidence and formulating recommendations, as shown in the table below (adapted from Schünemann et al 2006[107]).

Table 23. Advantages and inconveniences of having a common approach to assess the quality of evidence.

In favour of a common approach	Against a common approach
• Less demanding approaches may lead to false positive conclusions • Would minimise the selection of approaches based on vested interests • Would avoid confusion between different approaches • Might promote more and better research	• An inappropriate common system for some kinds of questions might lead to false negative conclusions • False negative conclusions from a common approach may lead to programmatic restrictions • Interventions which are not based on randomised trials might no be evaluated • A common approach may not discriminate within the range of evidence that is appropriate for health systems level interventions • A common approach may become overly complex

5.1.2 Outputs: presentation of outcomes profiles

Once the quality of the evidence on the outcomes has been assessed, the information will be used by the team developing guidance. This step requires more attention than the others because the team of guidance developers is formed by people with different backgrounds and disciplines who need to understand the implications of this information and, additionally, the way information is presented may heavily influence its interpretation and what is inferred from it.

An outcomes profile, which is proposed here (Figure 11), has the following characteristics:

- incorporates all outcomes studied across all interventions of interest in a single body of information;
- shows in a graphical format both quantitative data (e.g. odds ratios) and features of interventions and outcomes;
- transcribes the GRADE criteria in the table;
- incorporates additional details, such as participants, health system setting and context;
- it combines the following information items in a single bubble graphic:
    - direction of the effect (benefit or harm): bubbles above or below the middle horizontal line;
    - size of the effect: distance between the middle horizontal line and the bubble;
    - importance of the outcome: from left (less important) to right (more important); in this figure the order is illustrative and would be decided by the guidance steering group;
    - precision: vertical lines in each bubble representing the confidence intervals (vertical lines touching the middle horizontal line means that the effect was not statistically significant);
    - quality of the evidence: size of the bubbles;
    - intervention where the outcome was tested: colour of the bubbles (this allows to compare outcomes by interventions and looking at patterns in the distribution of bubbles corresponding to the same intervention);
    - whether an outcome of interest was not tested (empty bubble);
- it can be adapted by adding or deleting columns in the table or adding or deleting outcomes;
- it takes a single page.

Figure 11. Outcomes profile.

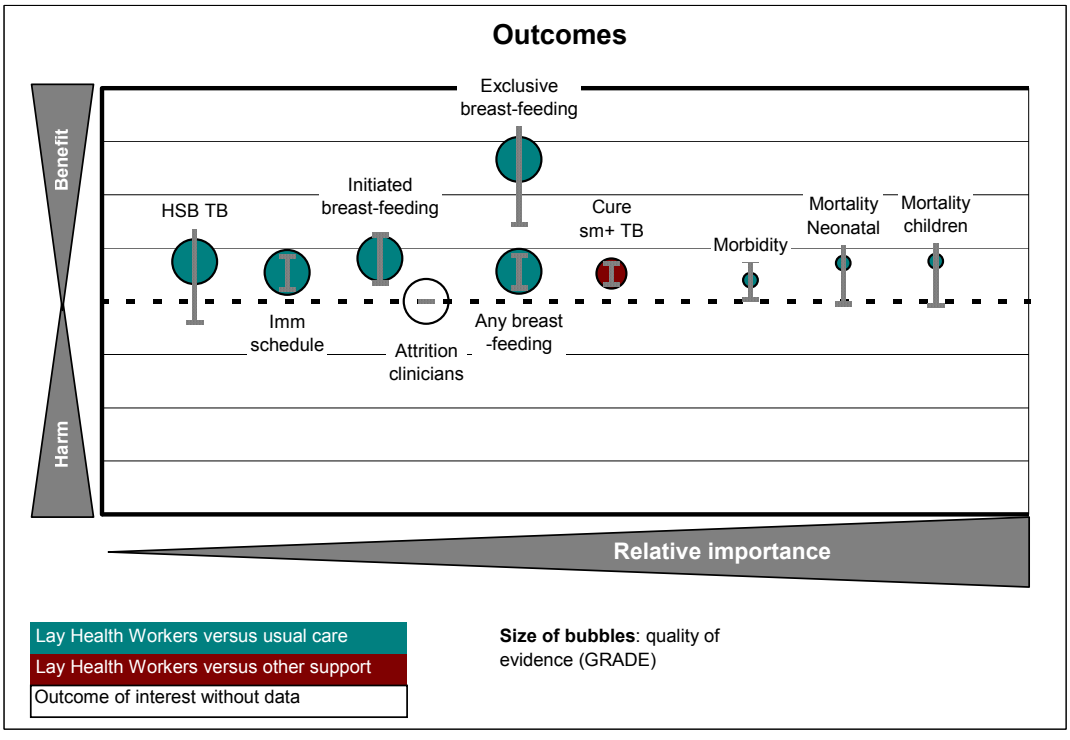
OUTCOMES PROFILE

Topic | **Use of Lay Health Workers for health service delivery**

	Intervention / Outcome	Importance	Data				Features	GRADE CRITERIA											
							Participants (equity)	Co-interventions	Health system	Geography	Magnitude	Confounding	Dose-response	Publication bias	Risk of bias	Imprecision	Inconsistency	Indirectness	Overall quality
	LHW vs. Usual care (Lewin 2010)																		
1	Imm schedule	2	**1.23**	1.09	1.38	RR	2/1											Mod	SR
2	Initiated breast-feeding	3	**1.36**	1.14	1.61	RR												Mod	Single study
3	Any breast -feeding	5	**1.24**	1.10	1.39	RR												Mod	
4	Exclusive breast-feeding	5	**2.78**	1.74	4.44	RR												Mod	
5	Mortality children	9	**0.75**	0.55	1.03	RR												vLow	
6	Mortality Neonatal	8	**0.76**	0.57	1.02	RR												vLow	
7	Morbidity	7	**0.86**	0.75	0.99	RR												vLow	
8	HSB TB	1	**1.33**	0.86	2.08	RR												Mod	
	LHW vs. Other support (Lewin 2010)																		
9	Cure sm+ TB	6	**1.22**	1.13	1.31	RR												Low	
	No evidence																		
10	Attrition clinicians	4				-												Mod	

RR: risk ratios; Mod: moderate; vLow: very low.

Shadowed areas contain the data used for the graphic below.



(See additional information sheet)

Since the contents of the profile are so succinct that the presentation fits onto a single page, it has to be accompanied by a brief information sheet of the definitions and explanations of the terms in the profile, as well as a brief overall interpretation.

The example in Figure 11 is drawn from the Lay Health Workers systematic review[108] with some dummy data for the sake of illustration. The table shows the odds ratios and their confidence intervals for nine outcomes, plus one outcome of interest which was not assessed. There are two comparisons: lay health workers versus usual care (in green colour) and lay health workers against other support for adherence (in brown colour).It could eventually contain other interventions as well. The columns to the right of the odds ratios show key features (participants, health system setting, context and other features) and the GRADE scores.

The graphic shows one bubble per outcome. Since outcomes can be measured in different ways (e.g. odds ratios, changes in proportions), there is no scale. Data is converted to make it qualitatively comparable. The outcome which had the greatest effect was 'exclusive breast-feeding'. Quality of all outcomes was moderate, except for the cure of smear positive tuberculosis patients (low quality, hence the smaller size of the bubble) and for the health status outcomes (very low), which were considered to be the most important outcomes. Although the effect on neonatal or child mortality is not statistically significant it is consistent with the other benefits which are statistically significant. This indicates that it is necessary to look at the whole picture not just on isolated significance levels particularly where the confidence intervals are wide and the statistical power relatively low. The empty bubble with the label 'attrition clinicians' indicates that this outcome was not assessed.

5.2 Evidence on implementation issues

Implementation issues are essential for issuing recommendations. It is useless to recommend an intervention with beneficial and large effects if it cannot be implemented due to technical, financial, political or other issues.

Evidence on implementation issues should be searched for, assessed and presented, as with evidence on outcomes. Where research evidence is not available, colloquial evidence can be taken into account at the stage of the deliberative process (see 6.2.2). When uncertainty is very prominent or if issues are greatly context-dependent, at least possible scenarios can be anticipated and discussed along the recommendations.

5.2.1 Approach: assessment of implementation issues

There are two tasks to be carried out:

1. to decide on the implementation issues that are relevant;
2. to assess and present the evidence in relation to these issues.

It is difficult to have a comprehensive list of issues[101,76]. In some cases, the factors considered are very few and complex; in others, there are many and not all of them are relevant in every situation. The ways they are formulated differ and may combine features of outcomes and interventions in the same category.

There are examples in the literature of how implementation issues have been addressed. A useful starting point could be to classify implementation into 'transferability' and 'applicability issues'[74]. Transferability issues include: baseline prevalence of the problem, characteristics of the population or system, and the capacity to implement the interventions; applicability issues: the political environment, social acceptability, cultural adaptability, resource implications and organisational structure and skills of local interventionists. Implementation issues could also be addressed under the perspective of barriers to access at different levels: community, household, health services,

health sector policy, strategic management, public policies across sectors and environment[109]. Alternatively, another framework identified constructs related to the intervention (e.g. evidence strength and quality), patient needs and resources, culture, leadership engagement and to individual characteristics to processes (e.g. plan, evaluate, and reflect. None of these models seems to have been compared to the other in terms of their capacity to better inform research and policy[110]. Whatever model is selected, issues need to be systematically analysed to complement the evidence on the effects of interventions (see also section 2.1.1).

ExpandNet is a "global network that seeks to promote equitable access to quality care by ensuring the benefits of successful health innovations are expanded to reach more people, more quickly and more sustainable"[111]. It provides a systematic approach to implementation issues. ExpandNet is about scaling up, defined as: "deliberate efforts to increase the impact of successfully tested health innovations so as to benefit more people and to foster policy and programme development on a lasting basis"[111]. Another term used, 'implementation', is defined as "to complete, perform, carry into effect (a contract, agreement, etc.); to fulfil (an engagement or promise)"[100].

The ExpandNet approach is suitable for developing guidance because it is comprehensive and includes key components of the guidance development processes[111]:

- it is about 'innovations': which is a 'set of interventions', including managerial processes for successful implementation;
- it considers different stages of innovation: pilot, demonstration, experimental or initial implementation stages;
- it entails a guided process;
- it takes into account institutional issues to ensure sustainability.

The ExpandNet framework for scaling up, adapted for the purpose of guidance development, includes the following components (in parenthesis the corresponding step in the ExpandNet guiding document):

1. the intervention (or innovation) itself (step 1): describes issues about the outcomes and about the interventions;
2. the implementers or facilitators (step 4);
3. the users of the interventions (step 2);
4. the context (step 3);
5. the scaling up strategies (steps 5 to 8).

Outcomes features are best addressed with the GRADE approach and they have already been presented in the previous sub-section. Points 2 to 4 are detailed in the second part of the table (orange colour in Table 19). The implementation menu of items considered in the ExpandNet approach is listed in the table below:

Table 24. Implementation items.

Category	Item
Intervention as tested	Costs
	Appropriateness
	Ease
	Testable
Requirements: facilitators	Credibility
	Skills
	Experience
	Size of team
	Resources
	Stability
Requirements: users organisation	Priority
	Capacity
	Timing
Requirements: context	Political
	Socio-economic
	Cultural
	Administrative
	Other sectors
	Donors
	Other stakeholders
	People's rights

These items are contextual-dependent, and therefore, there cannot be blue-prints with universal validity. Some of the features of interventions can be flexible and can be adapted without the risk of losing much of the effectiveness and which are key to success. What is relevant here is to indicate how sensitive the set of interventions of interest are to those factors. A robust intervention would be one that is relatively independent of them; in other words, what are the effects of those factors in deploying the intervention or set of interventions. In this way, policy will be informed not only on theoretical grounds but with evidence framed by the implementation issues (e.g. the capacity needed to actually deploy an intervention).

Evidence on those factors may come from systematic reviews containing descriptions about implementation issues that played a role when testing the intervention of interest, primary research evidence or colloquial evidence. The latter two are the most common sources.

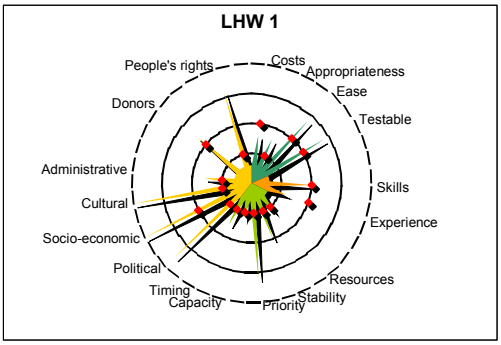
Evidence on implementation issues has to be assessed as has been done for the evidence of the interventions of interest in the main outcomes. If quantitative outcomes exist, they have to be assessed using a systematic and validated approach (e.g. GRADE). When quantitative outcomes are not available, but only qualitative research or colloquial evidence can be produced, evidence has to be made explicit as well by including the sources and any consideration useful for the interpretation.

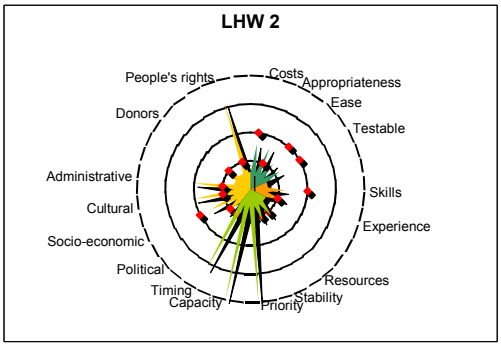
Since it can be anticipated that the quality of evidence on implementation issues will be relatively low in most cases and heterogeneous in relation to the sources of evidence implementation issues can be scored using an explicit scale[74]. The score would represent to which degree each factor influences implementation of an intervention. An additional quality of evidence score would accompany these scores, enabling to score as well ‘colloquial’ evidence.

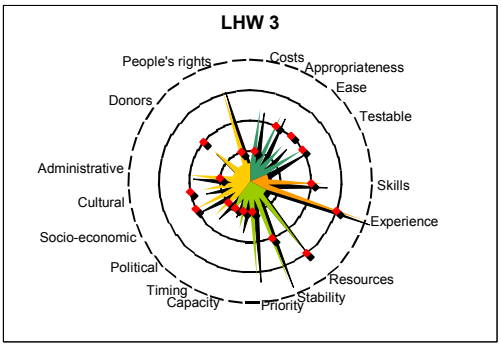
Figure 12. Interventions profile.

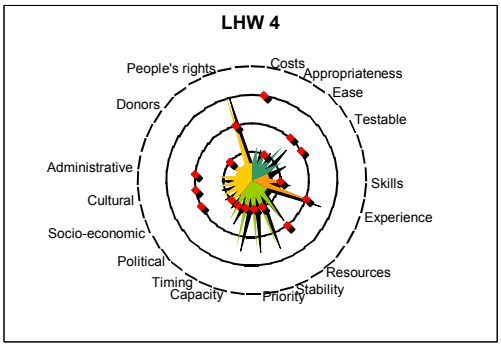
INTERVENTIONS PROFILE

Intervention / comparison:		LHW 1		LHW 2		LHW 3		LHW 4	
		Score	Quality of evidence(*)	Score	Quality of evidence(*)	Score	Quality of evidence(*)	Score	Quality of evidence(*)
Intervention as tested	Costs	4	2	4	2	6	1	3	3
	Appropriateness	4	1	4	1	6	2	3	1
	Ease	7	2	3	2	4	2	4	2
	Testable	7	2	3	2	5	2	3	2
Requirements: facilitators	Credibility								
	Skills	5	2	3	2	6	2	3	1
	Experience	3	2	3	1	10	3	6	2
	Size of team								
	Resources	3	1	3	1	7	3	4	2
	Stability	5	1	3	1	9	2	6	1
Requirements: users organisation	Priority	8	1	10	1	8	1	6	1
	Capacity	3	1	10	1	4	1	6	1
	Timing	4	1	8	1	3	1	4	1
Requirements: context	Political	9	1	3	1	4	1	3	1
	Socio-economic	10	2	4	2	5	2	3	2
	Cultural	10	1	5	1	4	2	3	2
	Administrative	4	1	5	1	5	1	3	2
	Other sectors								
	Donors	6	2	3	1	4	2	3	1
	Other stakeholders								
	People's rights	8	1	8	1	8	1	8	2









(*) 1: colloquial evidence; 2: low quality; 3: moderate quality; 4: high quality. *(Dummy data)*

It is also anticipated that these scores will entail much debate within the different groups inovolved in guidance development. However, this does not preclude that a small team of researchers and selected stakeholders carry out an initial assessment.

5.2.2 Outputs: presentation of interventions profiles

Implementation issues depend on each intervention or the way to scale up an intervention, and are therefore presented for each intervention of interest that may take part in the menu of recommendations. A table showing the score of the degree of potential influence of each factor with the quality of evidence is shown in the interventions profile. The table is accompanied by one spider graph for each intervention representing the scores and the quality assessments (Figure 12).

The interventions profile has the following characteristics:

- incorporates implementation issues across all interventions of interest in a single body of information;
- includes both scores and quality assessments;
- allows easy comparison of scores and quality assessments;
- it combines the following information items in a single spider graph:
    - category of the implementation issues (different colours);
    - score given to each issue (size of the spikes);
    - quality assessment of each score (bullet);
- it can be adapted by adding or deleting columns and rows;
- it takes a single page.

Figure 12 shows the examples of four interventions around lay health workers (LHW) using dummy data (e.g. LHW supported with supervision, LHW with special training, LHW receiving incentives, and so on). The numerical scores for each feature and intervention indicate to which extent the feature can be a barrier for implementation (for example, costs in intervention LHW3 seem to be much higher than in the other interventions). Scores are graphically represented in the spikes of the spider graphs below. Looking at the four interventions in the spider graphs, it can be seen that barriers are much more prominent in interventions LHW1 and LHW3 (large spikes). Some of the issues seem prominent for all interventions (e.g. ‘Peoples’ rights, ‘priority’), and some were not assessed (e.g. ‘credibility’, ‘size of the team’). The quality of evidence for each feature is represented by the red diamonds: the closer to the centre the less quality of evidence.

The issues could also be grouped or ordered by the difficulty in addressing them. In this way, two different types of information would be combined for each issue: how important is it to influence implementation and how feasible is it to act on it.

5.3 Additional considerations

The next step (“How to translate evidence into guidance and recommendations”) cannot take place unless this step is properly completed. It is essential to plan for it from the beginning, identifying the key personnel who will be in charge of the assessments and its presentation.

- Team:
    - Researchers to assess the quality of evidence.
    - Researchers and people with implementation experience to assess the implementation issues.
    - Communicators, in collaboration to the former ones, to prepare the profiles for the team using the evidence to issue recommendations.
- Resources
    - GRADE manual and software
    - Spreadsheet software or another software for graphical presentations
    - Handbook templates
- Timeframe

- Time will depend on the amount of evidence to assess, on its complexity, on the need to retrieve primary research from the systematic reviews retrieved, and on the team available to do so. Two months have been allocated in the template for the timeline (see Chapter 10).

Research topic 4. How the format of evidence summaries influence panellists judgements.

The way in which information is presented influences how it is understood and interpreted and, eventually, the judgments made about it. Furthermore, panellists involved in the deliberative process of issuing guidance have a wide range of backgrounds, including researchers, policy makers and lay people.

- How can evidence profiles be presented in a unique and understandable way to a wide range of stakeholders involved in issuing recommendations?
- How can different formats influence the judgements around the evidence presented?

6 How to translate evidence into guidance and recommendations

Mon dessin ne représentait pas un chapeau. Il représentait un serpent boa qui digérait un éléphant. J'ai alors dessiné l'intérieur du serpent boa, afin que les grandes personnes puissent comprendre.
[My draw did not show a hat. It showed a boa snake digesting an elephant. I, then, drew the inside of the snake so that important people could understand]
(Antoine de Saint-Exupéry. Le Petit Prince)

This is the most critical step in guidance development because actual guidance is produced in the form of recommendations. Recommendations are the result of judgements, and not merely a transcription or automated combination of evidence. Judgements respond to the logic of weighing what is beneficial 'against' what can be done. If something is not beneficial or cannot be done there is no point in recommending it. 'Beneficial' means improving the current situation described in the problem framed for guidance development; i.e. it can be reasonably foreseen that the outcomes that describe the status of the problem will improve. 'Can be done' means that the minimal requirements for an intervention or group of interventions to be implemented are met, including technical, political, societal or other issues that merit consideration. Since evidence on health systems outcomes and interventions will never be extensive, consistent and of the best quality –not even in the most optimal conceivable situation– there is always a grey zone where judgments will have to elucidate whether the benefits of the outcomes outweigh the burden of implementing the intervention or set of interventions.

6.1 Principles for issuing recommendations

A series of principles, taken from clinical guidance development, can be useful to consider before addressing the approaches for issuing recommendations (adapted from NICE[112]):

1. An intervention should not be recommended if there is no evidence, or not enough evidence, on which to make a clear decision. However, an intervention can be recommended in Monitoring and Evaluation (M&E) mode if this will provide more information about its effectiveness, safety or cost.
2. Those developing guidance must take into account the relative costs and benefits of interventions (their 'cost effectiveness') when deciding whether or not to recommend them.
3. Decisions about whether to recommend interventions should not be based on evidence of their relative costs and benefits alone. Other factors when developing guidance must be considered, including resource use and the need to distribute health resources in the fairest way within society as a whole.
4. Economic evaluations embedded in recommendations have to be explained from the point of view of their methodology and the rationale linking the results of the evaluation with the recommendations.
5. Although particular providers and users will expect to receive interventions responsive to their needs, this should not impose a requirement on guidance development to recommend interventions that are not effective, or are not cost effective enough to provide the best value to providers and users as a whole.
6. The institution issuing guidance should consider and respond to comments it receives about its draft guidance, and make changes where appropriate. Although the institution must use their own judgement to ensure that what it recommends is cost effective and takes account of the need to distribute health resources in the fairest way within society as a whole.
7. Recommendations can be restricted to a particular group of people within the population (for example, people under or over a certain age, or women only), but only in certain circumstances. There must be clear evidence about the increased effectiveness of the

intervention in this subgroup, or other reasons relating to fairness for society as a whole, or a legal requirement to act in this way.

8. When choosing guidance topics, developing guidance and supporting those who put its guidance into practice, the institution developing guidance should actively consider reducing health inequalities including those associated with sex, age, race, disability and socioeconomic status.

6.2 Approaches

Issuing recommendations involves combining research evidence (1) on the effects of interventions and (2) on implementation issues, as well as colloquial evidence, within a deliberative process that includes a panel of relevant stakeholders. The following two sections address how to combine evidence and the setting up of the deliberative process for issuing recommendations.

6.2.1 Combining evidence and grading recommendations

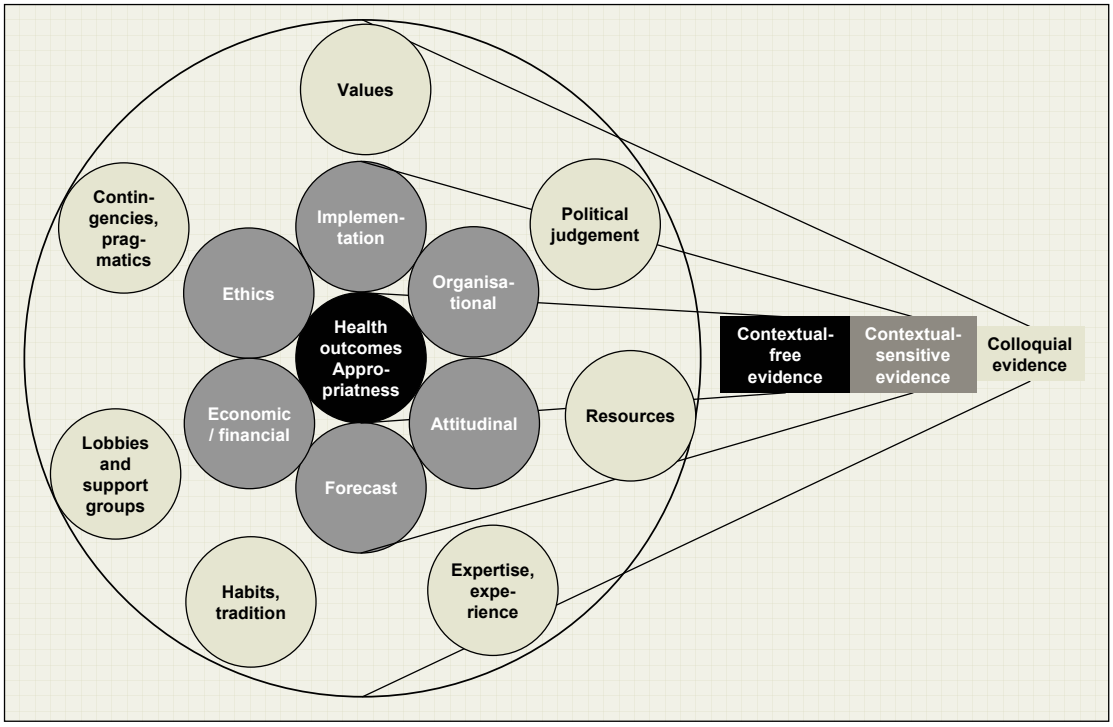
The goal of guidance development is not the creation of a 'prescription' but the development of context-sensitive guidance[53]. As shown in former chapters, in health systems guidance development, both evidence on the effects of interventions and on implementation issues is specifically searched, appraised and presented. The latter is not only retrieved where available from research synthesis reports of effects on interventions, but explicitly searched for by looking at reviews of study designs that deal with implementation issues (e.g. qualitative research). Neither of these two bodies of evidence will suffice to issue recommendations. Interventions may be very effective but costly and ultimately unfeasible; while other interventions may be easily implemented but may lack solid evidence to recommend them.

Furthermore, colloquial evidence from expert opinions or stakeholder views plays a role in issuing recommendations, as contextual issues cannot be totally grasped from research and its role cannot be avoided. Table 25 briefly presents these different types of evidence and how they are handled in guidance development. All are ultimately combined in the deliberative process set up for issuing recommendations (Figure 13 represents the same concepts in a graphical form).

Table 25. Types of evidence and their inclusion in guidance development.

Type of evidence	On what?	Source	Handling in guidance development
Research evidence	Effects of health systems interventions on outcomes	Research: experimental, quasi-experimental	Quality assessed; systematic incorporation into grading of recommendations
	Implementation issues	Research: most commonly non-experimental, qualitative	
Colloquial evidence	Contextual, political and pragmatic issues	Expertise, views and realities of stakeholders	Incorporation into grading of recommendations in the deliberative process

Figure 13. Types of evidence for context-sensitive health systems guidance.



Adapted from Lomas 2005[53].

Evidence on the effects of interventions and implementation issues can be systematically processed using tools that grade the strength of recommendations (see the following sections); while colloquial evidence is used to tailor the recommendations to the specific context where they will be eventually implemented.

In the deliberative process panellists can use several approaches to interpret and further elaborate on the evidence provided to them. Panellists have to agree on:

1. the balance between desirable and undesirable effects;
2. to which extent interventions are feasible (this includes costs);
3. how the balance between benefits and harms compares with the feasibility of implementation.

The next three sections address the issuing of recommendations from the perspective of a single intervention (using permutation matrix), comparing multiple interventions or options (balance sheets) and proposing decision aids to assist panellists when evidence is incomplete.

a. Grading recommendations

The relevance of grading recommendations is not the grading itself, but its implications. Grading recommendations reflects “the extent to which we can be confident that the desirable effects of an intervention outweigh the undesirable effects”[113]. Strong recommendations imply that they can be adopted in most circumstances while weak recommendations will need additional considerations (e.g. decision aids).

GRADE proposes a dichotomous approach: strong and weak recommendations. This depends on four factors:

1. Balance between desirable and undesirable effects: this balance has to do with the effects (desirable and undesirable) of the interventions. The larger the balance favouring desirable effects, the stronger the recommendations will be;
2. Quality of evidence: the quality of evidence describes the degree of confidence that can be placed in the measured effects. High quality evidence suggests that the measure of the effects (large or small, desirable or undesirable) is reliable; and vice-versa when the quality of evidence is low.
3. Values and preferences: this factor 'qualifies' to which extent desirable effects are actually valued and undesirable effects are seen as 'detrimental';
4. Costs (resource allocation): costs are highly dependant on context and even on the timing when the recommendation is adopted, as costs vary in different countries, settings and years.

These criteria can be complemented by other criteria[113] such as the importance of the effects (e.g. the effects may be seen as irrelevant even if large) or the implications of recommendations (e.g. it can affect a component of the health systems or a sector of the population that may not be perceived as a priority). For example, the grading of recommendations on increasing access to health workers in remote and rural areas also took into account feasibility[20] and the experiences and opinions of expert group members[20]. This guidance categorises recommendations as 'strong' and 'conditional' (i.e. 'very low' or 'low' quality of evidence, small magnitude of effects over a short period of time, more potentially negative effects, wide variability in values among stakeholders and significant variability between countries in the prerequisites for implementation).

Evidence on health systems interventions tends to be classified as of moderate or low quality at best, because of the rigor of health systems research and the types of study designs commonly used. Therefore, applying the GRADE criteria in a restrictive manner may lead to classifying all recommendations as 'weak' unless very strong additional arguments exist. On the other hand, for health systems guidance many additional factors apart from 'values and preference' and 'costs' need to be taken into account. Although this is considered in the GRADE tool, implementation issues in relation to health systems interventions tend to be more complex and numerous. Implementation issues are so prominent that evidence on them is explicitly searched for during guidance development (see the list of implementation issues in Table 24, page 69).

A 'permutation matrix'[114] can complement the GRADE tool by offering a more continuous scenario where additional categories of 'strength' of recommendations are made explicit and implementation issues are made more relevant by plotting them against the effects of interventions. Table 26 shows a generic example of how this could work: in the rows, the balance between benefits and harms are classified in four categories; and in the columns the same is done with the feasibility of implementation (the number of categories in rows and columns could vary according to the preferences of the guidance developers). The balance of benefits and harms can somehow include the assessment on the quality of evidence; i.e. if the benefits are much larger than the harms but evidence is of low quality, the net benefit balance would be reduced, and vice-versa. The same would apply for the evidence on feasibility issues.

If the balance between benefits and harms favours harms (regardless of the quality of evidence), implementing the intervention or group of interventions would be discouraged. If the balance is unclear, the decision to implement would depend on the feasibility of the 'new' intervention or group of interventions compared with the status quo. Finally, if the

balance favours the benefits, the intervention would be recommended if it is feasible or if feasibility issues can be addressed.

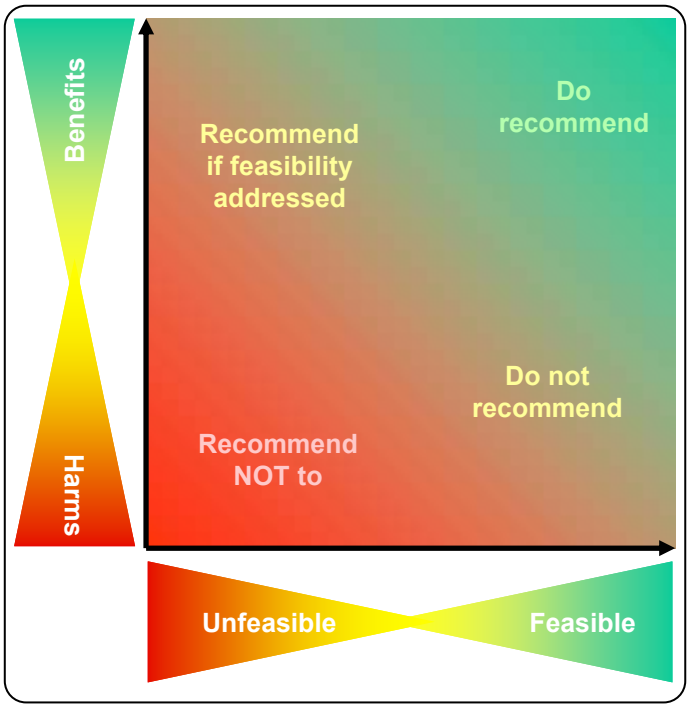
Figure 18 represents the same logic but uses a continuum without categories. Guidance developers may want to consider depicting interventions using pairs of data: one scoring the balance between benefits and harms (for the Y axis) and another for the implementation issues (X axis).

Table 26. 'Permutation matrix' to issue recommendations.

		Feasibility of implementation compared with current practices			
		Less	The same	More	Much more
Benefits compared with harms and with baseline outcomes	Much better	R?	R	RR	RR
	Better	R?	R	RR	RR
	The same	NR	?	R	R
	Worse	RN	RN	RN	RN

RR: strong recommendation; R: recommend; R?: uncertain whether to recommend; NR: do not recommend; RN: recommend not to do it. Colours are only for clarity.

Figure 14. 'Permutation blanket' to decide on recommendations.



b. Comparing interventions

It is anticipated that guidance on a given topic will produce evidence on a number of different interventions. Different options will have advantages and inconveniences.

Balance sheets[115] are used to represent benefits, harms, costs and by extension, other issues influencing the strength of recommendations, in a tabular form across different options[116]. Balance sheets can also (list[117] adapted for a health systems perspective):

- be used to assist decisions regarding specific contexts or health systems settings;
- show projected impacts of health system changes;
- become decision aids for policy makers and stakeholders.

The example below[118], related to the systematic review on lay health workers[108], illustrates a partial view of an online dynamic table where interventions are listed, and the effects and quality of evidence for different types of outcomes are shown in each column. This table gives, at a glance, a comparative impression of what works for different interventions and outcomes. Extra columns can be added to the right to collect explicit judgments on additional parameters, such as preferences and values or feasibility issues. The table also contains the quality of evidence assessed using the GRADE tool.

Balance sheets should show the outcomes of interest for guidance development, as defined in the scope of guidance and in framing the problem. Interestingly, if evidence on any of those outcomes is missing it would be reflected in the balance sheet, which will be very revealing when developing recommendations. If no alternative interventions exist, counterfactuals, either theoretical or reflecting the status quo, can be incorporated into the balance sheets.

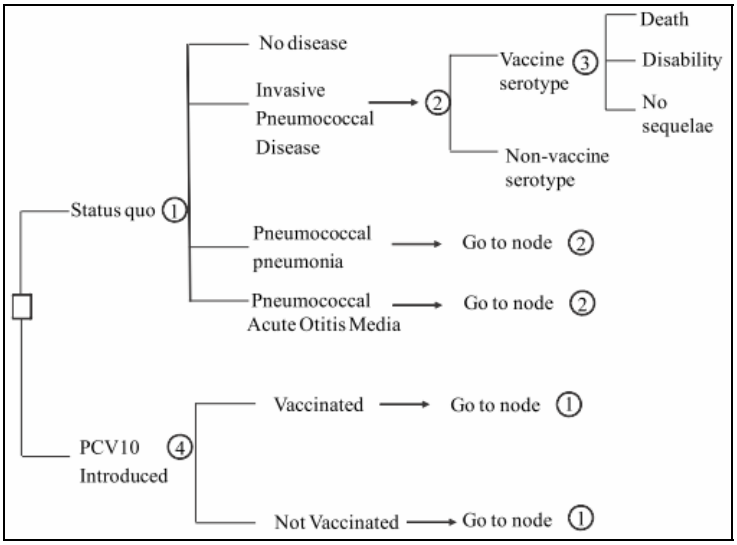
Table 27. Table cross-matching interventions and outcomes.

Lewin 2005	Intervention / question	Outcome	Access	Coverage	Quality of care	Morbimortality	Behaviour
Lewin 2005	Question: Should lay health workers be used for promting breastfeeding?	Breastfeeding initiation					RR 1.69 (0.91 to 3.12) [+++~]
Lewin 2005	Question: Should lay health workers be used for reducing mortality and morbidity?	Under-five morbidity				RR 0.86 (0.68 to 1.07) [+++~]	
Lewin 2005	Question: Should lay health workers be used for reducing mortality and morbidity?	Under-five mortality				RR 0.69 (0.51 to 0.94) [++++]	
Lewin 2005	Question: Should lay health workers be used for to promote immunisation uptake?	Immunisation uptake	RR 1.30 (1.14 to 1.48) [++++]				

c. Decision analysis aids

Other decision analysis aids[119] can be made available to the panel members in the deliberative process. For example, decision trees and/or mathematical modelling generate scenarios that can be constructed even if only incomplete evidence is available. These models are tested using sensitivity analyses that show how the final outcome of the decision tree changes when the assumptions change. This type of decision analysis has been widely used in clinical decision making. The advantages of these tools are that everything can be expressed in quantitative terms (e.g. values and preference may be assigned a quantitative 'weight') and that a single output is produced for each strategy tested. The main drawback is that if used recklessly it may lead to compelling but misleading conclusions supported by numerous assumptions. This approach would, however, make 'colloquial evidence' explicit and the assigned weight transparent. An example is shown in Figure 15[120].

Figure 15. Example of decision tree for the introduction of 10-valent pneumococcal conjugate vaccine.



6.2.2 The deliberative process for issuing recommendations

The aim of the deliberative process is to gather a panel of stakeholders to issue recommendations based on the available evidence, which has already been retrieved, appraised and properly presented (see preceding chapters and sections). This deliberative process does not aim to appraise evidence or to establish a policy dialogue to decide on the implementation of recommendations in specific settings, which will take place later on in the uptake of guidance by policy makers.

a. Setting up the deliberative process

The deliberative process is justified when[121]:

- Evidence from more than one discipline is involved
- Stakeholders have conflicting interests
- Evidence is uncertain or controversial in itself or because of contextual issues
- Issues of outcomes and costs go beyond the narrow scientific domain
- There are social and personal values involved, including equity and fairness
- There are issues of implementability and operational feasibility
- Wide public and professional "ownership" is desired

Developing health systems guidance easily meets all of the points above. The deliberative process involves several stakeholders with varied backgrounds, different degrees of involvement in guidance development and a diverse understanding of evidence, methodological issues and reality. This diversity is valuable to the deliberative process, however it is also challenging to conduct the process in an efficient manner that leads to sound outputs (the recommendations), consistent with the work completed thus far. The role of the facilitator cannot be overemphasised. The facilitator has to be familiar with the guidance development process and proficient in managing varied groups of people to lead them to a common goal.

The deliberative process does not exclude decisions aids such as decision pathways, algorithms or other techniques. These techniques should be used within the deliberative process to structure discussions.

Some tips on running guideline development group meetings are listed below (adapted from WHO handbook[1]).

- Clearly lay out the scope of the meeting at the start, including:
    - what is expected from meeting participants;
    - what needs to be achieved during the meeting;
    - what can be done afterwards;
    - what follow-up will take place with meeting participants.
- Choose someone who is experienced in facilitating meeting processes to chair the meeting. Keep to the agenda and organise the work using a reasonable schedule.
- Aim for a reasonable number of members in the panel. Some members may act as representatives of a small group of participants who can act as observers. Have a balanced expertise, including:
    - guidance developer experts;
    - topics experts:
    - health care providers, with experience / understanding of the issues affecting disadvantaged health care providers;
    - health care managers, with experience / understanding of health systems in resource constraint settings;
    - health policy makers, with experience / understanding of the health sector in low- and middle- income countries;
    - lay end-users of services, with experience / understanding of the issues affecting disadvantaged populations;
    - health technology experts and economists.
- Provide a clear set of documentation, consistently presented, well organised and sized to the purpose of the deliberative process.
- For issuing recommendations:
    - distribute the evidence materials before the meeting;
    - at the meeting, present draft recommendations that have been prepared by the guideline steering group (meeting participants will comment on these and refine them).
- Management of conflict of interests:
    - collect declarations of interest well before the meeting so that there is enough time to intervene if necessary (e.g. if any invited participant needs to be excluded due to major conflicts or because too many participants with potential conflicts of interest are present);
    - at the meeting, have each participant verbally report potential conflicts of interest; present a draft statement of declared conflicts of interest that will be refined and that will be presented again at the end of meeting for sign-off.
- Have someone write and project important decisions on a screen (i.e. research questions, recommendations)

All panel members should have access to the same information; namely:

- Plan an agenda of the whole deliberative process with objectives and a description of the outputs.
- Participant profiles, including their affiliations, to explicate who is who.
- A brief methodological guide containing the key concept definitions to be used during the deliberative process and to assist panellists in discussing and issuing recommendations (see an example for patient / carer groups contributing to technology appraisals[122]).
- Evidence profiles on (a) the effects of interventions and (b) implementation issues accompanied by background evidence materials.
- An initial proposal with graded recommendations.

b. Modalities of the deliberative process

The deliberative process, like any other approach used to develop guidance, has to be structured, systematic and transparent. Furthermore, there is evidence suggesting that different approaches may lead to different results. The term 'deliberative process' is used to designate this phase of guidance development and does not point at any specific method of conducting it.

There is some evidence suggesting that formal methods work better or at least equal to informal consensus[123]. It is suggested to use formal methods, as they are more likely to be transparent and easier to document and replicate. A recent Cochrane review assessing the effects of methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information materials found no trials involving clinical guidance development[124]. Although there seems to be no evidence on the effects of consumer involvement, other reasons to warrant their addition to the deliberative process include the convenience to incorporate their views in the process of issuing recommendations and to promote ownership. Other more fundamental ethical reasons may exist as well. This can be done rather formally when consumers are actively invited to participate in the guidance development process[125].

There are many methods to reach consensus, to understand decision making processes and many psychological and cognitive factors influencing how individuals make judgements. Some examples of methods to reach a consensus are illustrated in Table 28. These are suggestions from which guidance developers can draw ideas. More than one method can be used, depending on the specific scope of guidance being developed. For example, a study comparing nominal groups with the Delphi technique[126] suggested combining both to obtain optimal results (they used a large number of participants in both groups).

Table 28. Informal and formal consensus development methods.

Method	Description	Type of contact	Interaction structured
Informal	Self-explanatory	Any	No
Delphi	Participants are mailed questionnaires based on the cues identified; responses are collated and sent back for feed-back. If quantitative 'responses' are sought, they can be statistically combined.	Mailed questionnaires	Yes
Nominal Group Technique (NGT)	First, private writing of ideas which are then shared. Discussion of ideas. Voting for options. Judgments aggregated statically.	Face to face	Yes
NGT RAND version	Individual rating of questions or issues; collation and sharing of ratings; after discussion, second individual rating and agreement estimated using statistical methods.	Face to face	Yes
Consensus development conference	Open meeting of around 10 people. Evidence presented by external experts, who withdraw from deliberations.	Face to face	No
Staticised group	Aggregation of views of individual experts.	None	Not applicable
Social judgment analysis	Mapping of individual decision mechanisms which are then made explicit. More appropriate to understand why consensus is not reached.	Face to face	No
Structured discussion	Discussions based on a set of instructions.	Face to face	Yes
Glaser state-of-the-art	A core group drafts a 'position paper' which is peer-reviewed by members designated by the core group; several interactions till final format.	Face to face and indirect	Yes

Adapted from Murphy 1998[127] and Fink 1984[128]. There may be great variations in the way these methods are applied in different settings.

6.3 Outputs: recommendations

The outputs of the deliberative process are the recommendations. Recommendations have to be clearly formulated to easily infer what actions to take. The guidance to increase access to health workers in rural areas provides an excellent example of presenting recommendations. Based on this guidance, Table 29 transcribes some of the recommendations in a tabular form with slight adaptations (not all items appearing the in guidance document have been transcribed). For example, the absolute magnitude of effect is omitted since it would be integrated into the parameter describing the balance between benefits and harms; the quality of evidence is expressed in the same column as the balance between benefits and harms; the other parameters are all included under 'implementation issues' to emphasize that the sources of evidence actually differ from the evidence on the effects of interventions.

Some other changes could be introduced:

- recommendations could be sorted by status so that stronger recommendations are at the top of the table; or there could be one table for each status: strong recommendations and conditional recommendations;
- the status of recommendations could be further classified into:
  - strong or imperative[129]: recommendations that can be confidently issued;
  - scaling up: recommendations that are better applied on a small scale and only then scaled up. Close M&E would be advisable (see Chapter 7 for M&E);
  - innovation: recommendation that needs to be accompanied by implementation research;
  - experimental: recommendation that can be tested under experimental conditions.
- depict the quality of evidence of implementation issues, which may be 'colloquial';
- practical and key implementation considerations: who would implement the recommendation, what would be the specific requirements; institutional capacity; this could ideally take the form of 'in situation 1, A seems more desirable; in situation 2, B seems more desirable'. Situations could differ in relation to the characteristics of the problem, the type of health system or on wider contextual issues.
- possible health systems consequences that need to be considered while implementing;
- key policy issues anticipating policy makers and stakeholders barriers to implementation.

Implementation issues can be very detailed as shown in an example of guidelines for the introduction of a new vaccine[130]. Recommendations can also be complemented with additional materials such as case studies or illustrative examples where recommendations have already been implemented[130]. However, care should be taken not to overrule the evidence generated through comprehensive and well conducted systematic reviews with case studies or colloquial evidence.

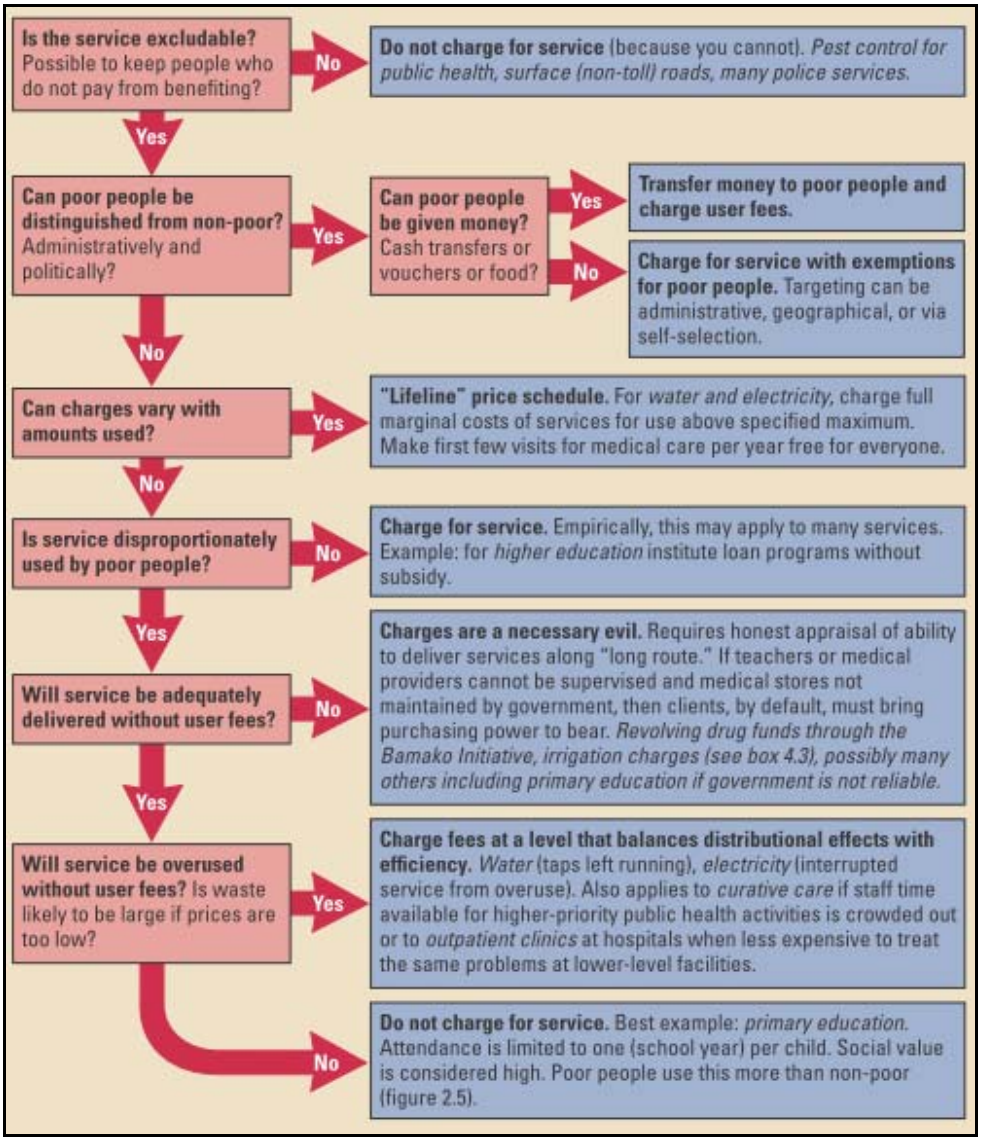
Another example of presenting recommendations is illustrated in Figure 16[131] showing a flow chart where implementation issues are used as decision points to select which intervention to implement.

Table 29. Example of presentation of recommendations.

Recommendation			Balance benefits-harms(+)	Implementation issues			Comments
#	Description	Status		Values and preferences(*)	Resource use(!)	Common issues(&)	
1	Use targeted admission policies to enrol students with a rural background in education programmes for various health disciplines in order to increase the likelihood of graduates choosing to practice in rural areas.	Strong	B+++	Yes	-	+	
2	Locate health professional schools, campuses and family medicine residency programmes outside of capitals and other major cities as graduates of these schools and programmes are more likely to work in rural areas.	Conditional	B++	Yes	+	-	
3	Expose undergraduate students of various health disciplines to rural community experiences and clinical rotations as these can have a positive influence on attracting and recruiting health workers to rural and remote areas.	Conditional	B+	No	+	-	
4	Revise undergraduate and postgraduate curricula to include rural health topics so as to enhance the competencies of health professionals working in rural areas and thereby increase their job satisfaction and retention.	Strong	B+	No	-	+	
5	Design continuing education and professional development programmes that meet the needs of rural health workers and that are accessible from where they live and work, so as to support their retention	Conditional	B++	No	+	-	
6	Introduce and regulate enhanced and safe scopes of practice in rural or remote areas to increase the potential for job satisfaction, thereby assisting recruitment and retention.	Conditional	BH+	Yes	-	-	
7	Introduce different types of health workers with appropriate training and regulation for rural practice in order to increase the number of health workers practicing in rural or remote areas	Conditional	BH++	No	-	-	
8	Ensure compulsory service requirements in rural and remote areas are accompanied with appropriate support and incentives so as to increase recruitment and subsequent retention of health professionals in those areas.	Conditional	BH++	Yes	+	-	

(+) B: benefits outweigh harms; H: harms outweigh benefits; BH: benefits and harms are balanced. Quality of evidence: ++++ high; +++ moderate; ++ low; +; very low. (*) Variability. (!) - less; + more. (&) Common feasibility factors across different countries: + likely; - unlikely.

Figure 16. Example of decision flow chart showing options for user fess policies.



Finally, an intermediate output of the deliberative process includes minutes of the interactions, records of key decisions, participants, and other records documenting the process.

7 How to monitor and evaluate (and research) recommendations

Those who cannot remember the past are condemned to repeat it. […]
In a moving world readaptation is the price of longevity
(George Santayana. The Life of reason I. Reason in common sense, Chapter XII)

The need to monitor and evaluate recommendations is justified by the following:

1. recognising that considerable funds go into health and development and that these funds will never suffice to cover the health needs of populations, evaluation (of programmes and policies) should be part of sound evidence informed policies[132];
2. recommendations on health systems are likely to be based on 'moderate' quality at best, and therefore carry a degree of uncertainty about the effects of the recommendations;
3. the main body of evidence supporting recommendations is typically generated in research settings, which are more controlled environments than real life situations;
4. in the area of aid and development, monitoring systems are seen as crucial (e.g. to achieve the MDG[133]); furthermore, the Paris Declaration on Aid Effectiveness mentions the capacity of countries to carry out M&E as one of the critical aspects for achieving development objectives[134,135] and it is part of the harmonization component[136];
5. monitoring and evaluation is an integral part of the planning cycle that guides the deployment of interventions, services and programmes[*];
6. finally, and maybe most importantly, by an ethical imperative based on the recognition that the lives of people are affected by the validity of policy judgements[137].

Monitoring and evaluation is a very wide field in itself with a great diversity of concepts and approaches[†], and it would be beyond the scope of this handbook to provide a full description of the approaches and methods currently being used. Instead, in the following sections, several aspects to be taken into account when planning for monitoring and evaluation will be proposed. In general, there is an immense body of knowledge in literature about the rationale, approaches, methods and tools for M&E. Many donors, international agencies (e.g. AusAID[138]), global health initiatives[139] and countries[140] have generated their own approaches to M&E, most of them consistent in the essentials but with great variations in the definitions and ways to undertake M&E.

'Monitoring' and 'evaluation' are terms that have been used ambiguously. Typically, monitoring has to do with the routine and repeated measurement of activity outputs carried out along a project or programme. Evaluation is usually done at the end of a project or programme (sometimes in the middle) to assess the degree of attainment of the objectives. The approaches and methods differ depending on what is measured and how often and on the availability of data or the difficulties in obtaining reliable data. In this chapter the term "Monitoring and Evaluation" is used as a single term to designate the assessment of a project, programme or recommendations in terms of its processes or outcomes in order to improve them and to allocate resources[141]. Therefore, M&E is conducted at national or sub-national levels where the actual options are implemented although synthesis of evaluations can be conducted at global level.

[*] Likewise, clinicians routinely monitor the health status of patients receiving treatments, even if treatments are known to be effective.

[†] 1. Cluster evaluation; 2. Country program evaluation; 3. Country assistance evaluation; 4. Ex-ante evaluation ; 5. Ex-post evaluation ; 6. External evaluation; 7. Formative evaluation; 8. Independent evaluation; 9. Internal evaluation; 10. Joint evaluation; 11. Meta-evaluation; 12. Mid-term evaluation; 13. Participatory evaluation; 14. Process evaluation; 15. Program evaluation; 16. Project evaluation; 17. Review; 18. Risk analysis; 19. Sector program evaluation; 20. Self-evaluation; 21. Summative evaluation; 22. Thematic evaluation.

7.1 Overview

A series of principles emerge from most of the M&E approaches being currently used in the areas of health, development and aid[*]. They are summarised in the table below:

Table 30. Monitoring and Evaluation principles, relevant to health systems guidance.

	Principle	Relevance for M&E of recommendations
1	Action oriented	The action mechanisms to respond to M&E findings are part of the M&E plan. The mere generation of knowledge is not enough justification for a full M&E plan.
2	Explanatory	M&E describes what goes right and wrong but should also offer plausible explanations of why activities were not conducted or objectives not achieved. A conceptual framework should be adopted from the start (e.g. inputs – process – outputs).
3	Participatory	Involvement of key stakeholders, including end users, or community members from the M&E planning stage up to the interpretation of findings and design of action points. A clear decision-making flow / structure needs to be in place and already functioning.
4	Timely	M&E is synchronized with the project / programme life cycle, donor or funding flows, and 'rules of the game' of the decisions making processes.
5	Integrated	M&E is integrated into general management and decision-making structures of an organisation, project or programme. 'Integration' means that M&E shares the resources and procedures of a wider working unit.
6	In scale	M&E should be kept in scale (in terms of workforce, time and funding) in relation to what is monitored and evaluated (it has been suggested that M&E accounts for around 5% to 10% of the running costs of a project or program[†]).
7	Harmonized	M&E requirements (e.g. data and data management facilities) are clearly defined and communication and information exchange is established with other projects, information systems or units in order to avoid duplication in data collection and to increase the availability and quality of data.
8	Tailored	M&E takes into account the nature of the projects or programs being evaluated (i.e. some programs may have added values in terms of their processes or inputs, proof of concepts, innovative methodological approaches). Interventions that have been extensively proven to be effective may not need the most rigorous evaluations.
9	User-friendly, feedback	M&E findings are shared with and understood by decision makers and other relevant stakeholders. Findings, either quantitative and/or qualitative, are presented in a user-friendly and simple way in order to ensure correct interpretation.
10	Transparent	M&E methods are unbiased and transparently used and described. This is ensured through a predefined analytical plan. The quality of data is also assessed.

[*] Based on the M&E experience of the Swiss TPH, evaluating projects and programmes and undertaking programmatic or institutional reviews.

[†] The Global Fund. Guidelines for the submission of an M&E plan for Global Fund grants. 2007.

7.1.1 Dimensions

There is a distinction between what to evaluate and where to look to evaluate. 'Where to look' has long since been defined by dissecting programmes or services into 'inputs', 'processes', 'outputs', 'outcomes' and 'impact'[142]. Despite the fact that each one of these components can be measured in itself, the 'depth' of the evaluation is given by the dimensions, which measure those components and their relations (e.g. inputs against outputs to assess 'efficiency').

Below, the DAC dimensions for evaluating development assistance[143] and aid programs or project activities are listed for reference. These are: 'relevance', 'effectiveness', 'efficiency', 'impact' and 'sustainability'. As defined in the OECD evaluation framework[144]:

Relevance: the extent to which the activity is suited to the priorities and policies of the target group, recipient and donor. Examples of evaluation questions include:

- To what extent are recommendations aligned with population needs and targeted health systems problems?
- Are recommendations and their outputs consistent with higher level policies?
- Are recommendations consistent with the intended impacts and effects?

Effectiveness: a measure of the extent to which an aid activity attains its objectives. Examples of evaluation questions include:

- To what extent recommendations have achieved their objectives?
- What were the major factors influencing the achievement or non-achievement of the objectives?

Efficiency: measures the outputs -qualitative and quantitative- in relation to the inputs. It is an economic term which signifies the resources that are used as the least costly possible in order to achieve the desired results. Examples of evaluation questions include:

- Were activities cost-efficient?
- Were objectives achieved on time?
- Was a given recommendation implemented in the most efficient way compared to its alternatives?

Impact: the positive and negative changes produced, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors. Examples of evaluation questions include:

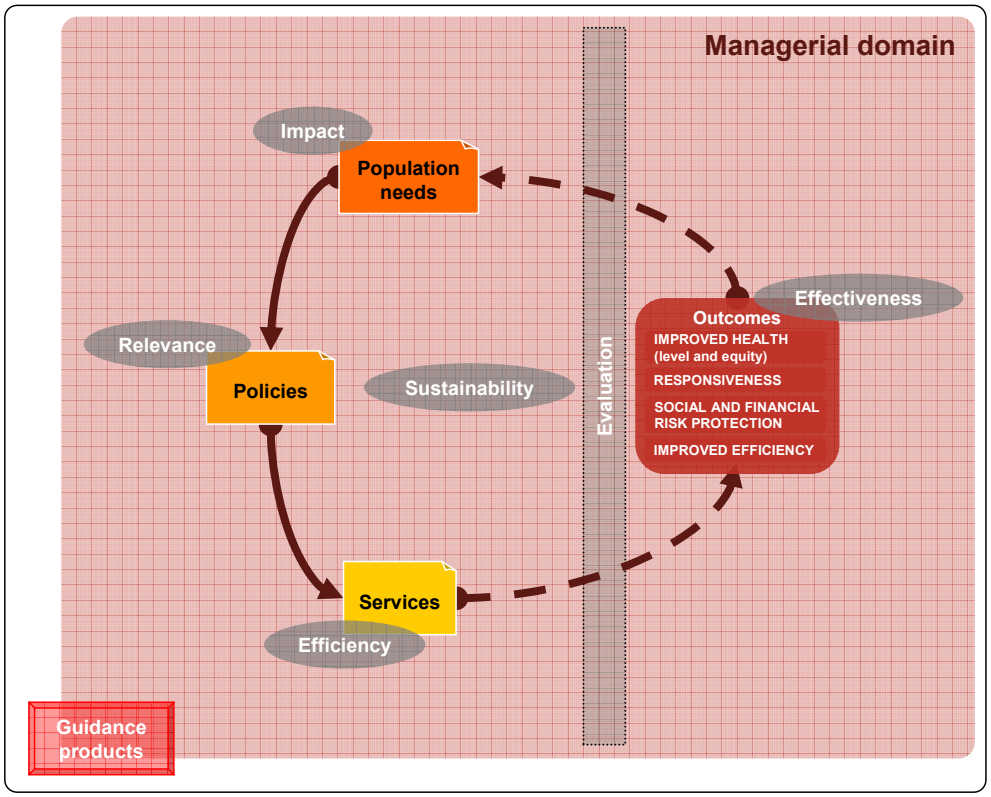
- What were the consequences of implementing the recommendation?
- What real difference has the recommendation made to the beneficiaries (health systems components and end users)?
- How many people were affected?

Sustainability: concerned with measuring whether the benefits of an activity are likely to continue under routine funding circumstances. It includes environmental as well as financial sustainability. Examples of evaluation questions include:

- To what extent do the benefits of a programme or project continue after any special funding ceases?
- What were the major factors that influenced the achievement or failure of sustainability of the programme or project?

These dimensions are depicted in Figure 17, which is an extract of the framework for health systems guidance (Figure A - 1).

Figure 17. Dimensions of M&E in the managerial domain of the health systems guidance framework.



M&E dimensions in the grey oval shapes. This is part of the analytical framework for health systems guidance. See Figure A - 1 for a complete representation of the framework.

There seems to be a consensus that M&E related to health systems has to address performance in terms of both health systems measures and population health outcomes[145]. This is consistent with the problem definition phase where both health systems components and the end-users of the system are taken into account.

7.1.2 Approaches

Every conceivable study design can provide findings for monitoring and evaluation, such as qualitative and quantitative study designs, observational and experimental studies, or of any other sort. The type of decisions to be drawn from the evaluation will dictate the methodology to be used. This will depend on other factors as well, such as additional existing knowledge and the resources available[146].

In essence, the selection of methods depends on what to measure in the evaluation (see section 7.1.1 Dimensions) and on the degree of certainty required in relation to the findings of the evaluation[146]. These are the two axis represented in Table 31.

The first axis defines the component or dimension to be evaluated: performance ('inputs', 'processes' and 'outputs', in the terms used in the preceding sections) or impact ('outcomes' and 'impact'). The first deals with 'activities' measured in relation to health care providers, health systems components or end users; while the second measures health status and behaviour in end users and in targeted health systems components.

The second axis is divided into 'adequacy', 'plausibility' and 'probability'. In the case of 'adequacy', the evaluation determines whether activities or observed changes have taken place according to a set of predefined criteria. In the case of 'plausibility', observed effects can be reasonably attributed to the intervention being implemented. This can be determined with statistical significance when 'probability' designs are used. At the bottom of Table 31 the types of studies corresponding to each degree of certainty are shown.

This classification is also helpful to understand the interests of different stakeholders: programme managers and decision makers may advocate for measurements of performance using 'adequacy' designs or some 'plausibility deigns', whereas scientists may be more interested in 'probability' designs to demonstrate impact. The interests of other stakeholders (e.g. funding agencies, partners) may fall in-between, most likely to demonstrate attribution using 'plausibility' designs.

Others have expressed the first axis in the framework Table 31 below in terms of 'monitoring' ('performance'), 'formative evaluation' ('performance' focusing on processes) and 'summative evaluation' ('impact'). 'Monitoring' has to do with activities and is usually undertaken internally by the same team who implements the programme or by managers directly involved in the programme. Evaluation, formative and summative, can be external to the project or programme being appraised or even independent when it is conducted by evaluators without any type of involvement in the project or programme. Table 32 provides some details on the timing and the information support for each type of evaluation.

Table 31. Dimensions and degree of certainty of Monitoring and Evaluation designs.

AXIS 1	AXIS 2: CERTAINTY		
	Adequacy	**Plausibility**	**Probability**
PERFORMANCE: Provision | Utilisation | Coverage			
What to measure	Programme activities		
In whom to measure	Implementation workers, programme recipients		
Inference that:	Activities as planned	Effect above and beyond the impact of non-programme influences	Intervention has a better effect than control (p)
IMPACT			
What to measure	Health and behavioural indicators		
In whom to measure	Programme recipients or target population, targeted health system components		
Inference that:	Observed change as expected	Effect above and beyond the impact of non-programme influences	Intervention has a better effect than control (p)
Comparison	Predefined adequacy criteria	"Opportunistic" or non randomised control	Randomised controlled group(s)
Typical study designs	Cross-sectional Longitudinal		
		Case-control	
		Longitudinal, controlled	

Adapted from Habicht 1999[146]. p: statistical p-value, or any other parameter of statistical inference.

Table 32. Monitoring and formative and summative evaluations.

Type of evaluation	Purpose ('action')	Agents	Timing	Information support
Monitoring (Performance monitoring)	Are the outputs produced? Are activities taking place? Are funds being used?	Self-evaluation / internal*	Continuous / periodical	Routine data
Formative evaluation (Implementation/Process evaluation)	Can performance be improved?	External	Single point before the end	Routine data ad hoc data
Summative evaluation (Outcome/Impact evaluation)	Are outcomes achieved?	External† and Independent (usually)	Single point at the end	Routine data ad hoc data

Finally, participatory M&E, as opposed to conventional M&E, is increasingly advocated. Participatory M&E involves local people, development agencies, and policy makers deciding together how progress should be measured and results acted upon[147]. Table 33 summarises the main differences.

Table 33. Conventional and participatory Monitoring and Evaluation.

	Conventional M&E	Participatory M&E
Who plans and manages the process	Senior managers, or outside experts	Local people, project staff, managers, and other stakeholders, often helped by a facilitator
Role of 'primary stakeholders' (the intended beneficiaries)	Provide information only	Design and adapt the methodology, collect and analyse data, share findings and link them to action
How success is measured	Externally-defined, mainly quantitative indicators	Internally-defined indicators, including more qualitative judgements
Approach	Predetermined	Adaptive

Adapted from IDS 2008[147].

'Participation' means opening up the design of the process to include those most directly affected and agreeing to analyse data together. This involves 'negotiation' to reach agreement about what will be monitored or evaluated and about the particularities of data sources and management, analyses, interpretation and actions to be taken. It is expected that this leads to 'learning' which translates into improvement. The participatory approach can be promising considering the complexity of the processes involved in implementing guidance, adapting and following recommendations and changing policies in health systems.

* Internal does not necessarily mean 'self-evaluation' but rather that evaluators belong to the same project or programme.

† External does not necessarily mean independent, since external evaluators can have vested interests in the results of the evaluation.

The guidance steering group may also consider the capacity to actually undertake M&E and what is required to do so, based on the parameters outlined in the preceding sections. Some hints to bear in mind include:

- institutional articulation of M&E activities within organisations or programmes; allocation of resources, including staff, time and funding (e.g. M&E in the organisational chart, with specific budget allocation);
- relation to other relevant initiatives (e.g. quality assurance, organisational programmes and research);
- human capacity and infrastructure to collect, enter, store and analyse data, and to interpret and report it (e.g. designated staff, office space, time allocated in job descriptions);
- formal links with senior management (and managerial processes) and decision making bodies (and decision making processes); (e.g. data supporting board decisions);
- clear strategies to translate M&E findings into action (e.g. data presentation and dissemination, follow up of actionable items);
- links with staff learning and development initiatives (e.g. coaching, mentorship);
- local knowledge exchange and dissemination (e.g. with similar projects and programmes);
- specific M&E training (e.g. staff in planning units, management).

7.2 Implications for health systems guidance

7.2.1 Dimensions

The guidance steering group will need to consider what is to be measured. It may be helpful to take the problem defined at the start of the guidance development process as a reference. Ultimately, what needs to be measured is actually the extent to which the problem that triggered guidance has been successfully addressed. Table 34 applies the parameters used in the problem definition (see Table 2, page 11) to suggest how the M&E dimensions fit in the way the problem was framed, including the approaches that could be considered for M&E.

‘Effectiveness’ measures to which extent the recommended intervention(s) is/are effective in improving the health status of the population and the ‘health status’ of the system, because health system level interventions target components of the health system. ‘Impact’ has to do with end goals in terms of health status or determinants of health of the population.

The problem modelling establishes the causal links of events within the health system and between the system and the end-users. ‘Relevance’ describes to which extent recommended interventions actually address the problem under the perspectives of the wider health system and the end users. Therefore, it has a predominant ‘process’ evaluation component.

‘Efficiency’ can cover aspects that relate to the way recommended interventions are delivered and how the context influences operational issues. Evaluation of efficiency needs to ‘explain’ how interventions can take place and are eventually scaled up in the specific geographical, health systems and socio-economic contexts.

Finally, ‘sustainability’ should be able to establish the likelihood that the requisites to deliver the recommended interventions (e.g. funding, stakeholders’ involvement) in routine conditions can be met in the long run.

Table 34. Framing the problem for health system guidance and implications for monitoring and evaluation.

	Description	Dimensions	Approaches and methods
Problem parameters	Determinants of health and health status	Impact	Subjects / components in whom / which to measure outcomes
	In terms of health system	Effectiveness	
	Population and health systems events by which the problem is identified as such		Outcomes evaluation
Problem modelling	Establish a causal chain linking health systems problems with users' needs	Relevance	Inputs evaluation Process / explanatory evaluation
Context	Geographical context, stakeholders and other factors influencing people and health systems.	Efficiency	Explanatory evaluation
Opportunities	Establish what makes your institution and the present time favourable conditions to address the problem	Sustainability	Process / explanatory evaluation

See also Table 2. Framing the problem for health system guidance.

The data sources for M&E will depend on the specific dimensions to be measured. In section 2.1 (see Table 2) it was pointed out that the selection of outcomes was guided by the parameters used to define the problem. Therefore, there should be some degree of consistency between the data sources to describe the problem and the data sources to M&E the effects of recommended interventions into the parameters of the problem. This will depend as well on the type of M&E (Figure 18). For example, when framing the problem, routine data may have been used to estimate its parameters (e.g. the geographical distribution of health workers and their qualifications in the different tiers of the health system). However, if a robust M&E approach is needed (e.g. 'probability' designs, Table 31), additional data sources and methods to collect them may be needed, such as repeated cross-sectional staff surveys conducted in areas where the recommended interventions are implemented and in control areas (see section 2.1.1 Approach: framing the problem, parameters and models). Table 3 (page 14) shows relevant data sources for each type of information: determinants of health, inputs, outputs, outcomes and health status ('outcomes' and 'impact'). Additionally, Table 35 lists examples of sources of quantitative and qualitative data that can be used to suggest methods that support data collection for M&E.

Table 35. Possible sources of quantitative and qualitative data.

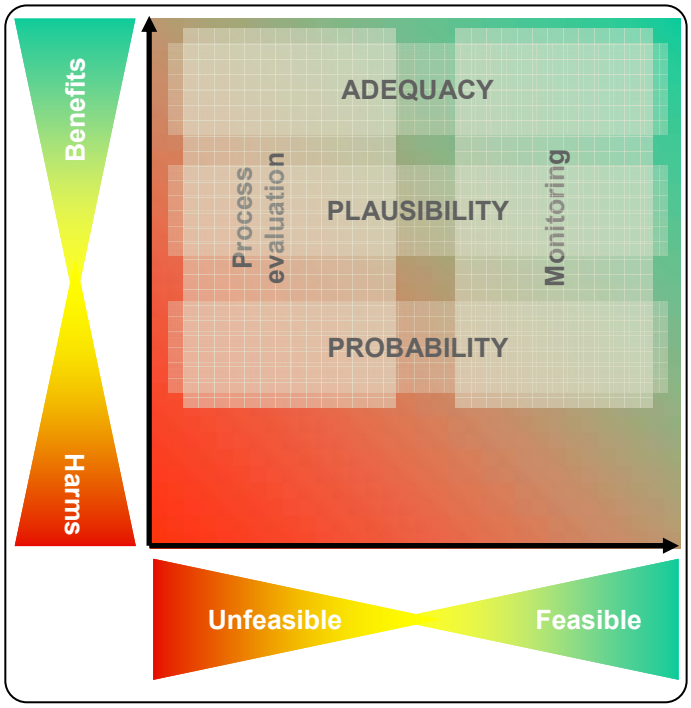
Sources	Quantitative	Qualitative
Household and community	• Surveys: e.g. DHS (Demographic and Health Survey), MICS (UNICEF-Multiple Indicator Survey). • Quality Assurance Samplings • Census based Household information systems	• Qualitative data may be recorded from focus group discussions or individual interviews to guide the quantitative methods or to help interpret the results.
Health facility	• Health facility surveys • Routine Health Management Information System • Supervision reports	• Focus group discussions with staff, exit-interviews, used in the same way as described above.
Health system	• Programme or service reviews • High level supervision reports	• Stakeholders analysis • Organisation analysis • SWOT analysis (see Unit 6)

7.2.2 Approaches

The guidance steering group will decide the most reasonable M&E approach for each recommendation or group of recommendations by considering how recommendations have been issued.. Figure 18 shows the 'permutation blanket' where interventions can be placed in the continuum 'benefit' (effects on outcomes) – 'feasibility' (implementation issues). The stronger the evidence on the benefits of the interventions the less demanding M&E of recommendations needs to be; hence, adequacy designs may suffice. However, if the evidence is weak or the effects are not dramatically beneficial, more robust designs may be needed to avoid causing unintended harmful effects without notice. On the other hand, if feasibility issues are judged to be minor and easily addressable, routine monitoring of the recommendations may be enough. However, more rigorous process evaluations may be needed when feasibility is challenging.

The lower part of the chart in the 'permutation blanket' (Table 14) is empty because when the effects of the interventions are harmful they are not recommended and M&E is unnecessary.

Figure 18. 'Permutation blanket' with evaluation approaches.



7.2.3 In practice: M&E and research

There are three related issues that may need clarification at this stage:

- 'Issuing recommendations' is based on the evidence provided in the guidance materials: evidence on the effects of interventions and evidence on implementation issues.
- M&E is not used to select recommendations, but rather to assess to which extent a selected recommendation produces the expected outputs and is carried out as reasonably foreseen.
- Finally, research is a methodological approach that can contribute to reduce evidence gaps used to issue recommendations and to provide data for M&E.

The 'permutation blanket' in Figure 18 suggests that under certain circumstances the limits between what constitutes M&E and research can be blurred. This should also be made clear when proposing approaches to M&E recommendations in order to generate appropriate expectations in relation to the burden of M&E activities, as well as when the interpretation and role of new evidence is generated by M&E or research. This is also an opportunity to use M&E as the framework to suggest ways to fill the evidence gaps identified when searching evidence that support recommendations (see section 4.2.3).

The guidance steering group may decide on different ways to present suggestions for M&E of recommendations; for example, providing specific guides for each recommendation or a generic framework for all of them (see Table 36[20]). Although it is advisable to follow existing and widely used M&E frameworks, the guidance steering group may very well choose alternative frameworks or to adapt to existing ones if needed, given the nature of the particular recommendations. For example, 'relevance', '(political) acceptability', 'affordability', 'effectiveness' and 'impact', have been suggested in the context of retention of health workers in rural and remote areas[20].

Table 36. Examples of questions and indicators for M&E of interventions to increase access to health workers in remote and rural areas through improved retention.

Stage	Questions to be asked	Indicators or measures of progress	Methods
Design	- Did the intervention respond to a documented need? - Is the choice of the intervention based on evidence or robust arguments?	- HRH situation analysis - HRH costed plan - Stocks and flows of health workers - Density of health workers in urban versus rural areas	- Labour market analysis - Demographic analysis (health workforce stocks and flows) - Surveys of intentions - Stakeholder analysis - Review of policy documents
Implementation	- Relevance: were the preferred choices of health workers for rural work identified? - Acceptability: have all stakeholders been engaged? - Affordability: have all sources of funds been identified and secured?	- Factors that motivate health workers to go to, stay in or leave rural areas - Stated preferences for rural job attributes - Stakeholders consultations and engagement - Budgets allocated for the proposed interventions	- Survey of intentions - Focused group discussions - Discrete choice experiments - Stakeholder analysis - Review of policy documents
Results	- Did attractiveness of profession/rural/remote areas improve? - Did recruitment of health workers in underserved areas improve?	- Total number of graduates of health professional schools - Preferences for rural/remote areas - Total number of health workers recruited to rural areas - Proportion of new graduates entering rural practice	- Analysis of registry data - Surveys, focus group discussions - Analysis of registry data or facility data

Table partially extracted from the 'Increasing access to health workers in remote and rural areas through improved retention[20].

8 How to disseminate and implement guidance materials

If everyone was listening, you know
There'd be a chance that we could save the show
(Supertramp. If everyone was listening)

Guidance dissemination refers to the process that makes guidance materials available to and ready to use for its intended end-users. Guidance implementation includes all those activities or strategies that lead to guidance actually being used in real life situations (measured, for example, by end-users adherence to guidance recommendations). Guidance dissemination precedes guidance implementation and it is a needed to enable the latter.

In this chapter of the handbook dissemination and implementation of guidance materials are addressed, although the actual dissemination and implementation of specific guidance produced may fall beyond the responsibilities of the guidance steering group.

8.1 Dissemination

The dissemination of guidance is the visible output of the whole process and what all this work leads to. What is disseminated is what will be eventually reviewed, shared and used. It is essential that dissemination correctly reflects the work done and that it encourages the use of guidance. The importance of dissemination cannot be overemphasised, especially when it appears less prominently than other aspects of evidence and guidance[148].

8.1.1 Pre-dissemination stage: the guidance 'master report'

The core of guidance is comprised of recommendations, which are supported by a body of evidence that has been systematically searched, retrieved and assessed for its quality. A comprehensive description of the scoping of guidance, methodological descriptions and documentation of the guidance development processes has to be compiled into a single, comprehensive report that will be used as reference for all the other materials that can be disseminated by other means. The principal end-users of health systems guidance are policy makers, and comprehensive and complicated reports' may not be the best means with which to communicate guidance.

A suggested structure of the guidance report is detailed in Table 37. This can be adapted to better suit the style and preferences of the organisation issuing recommendations. It would be advisable to use the same format and structure as a 'brand' of the organisation throughout all guidance.

The 'dashboard' is a single page containing the key technical information (evidence and recommendations) using a combination of text, tables and graphics to provide a synthetic view of guidance at a glance. The following sheet should detail the information related to the authorship of guidance, potential conflicts of interest, funding and the due date of the next update. The guidance summary, which is more extensive, should be no longer than three pages, and describe the background, scope, methodological aspects, evidence and recommendations. The other sections of the document are self-explanatory.

Table 37. Table of contents of the guidance 'master document'.

1 Dashboard page
2 Institutional and experts credentials page
3 Guidance summary
  3.1 Background
  3.2 Problem definition and scope of guidance
  3.3 Approaches and methods
  3.4 Evidence and recommendations
4 Background
5 Problem definition and scope of guidance
  5.1 Search for existing guidance and evidence
  5.2 Inclusion and exclusion criteria
6 Approaches and methods
  6.1 Search for existing guidance and evidence
  6.2 Inclusion and exclusion criteria
  6.3 Appraisal of the evidence
  6.4 Issuing recommendations
7 Evidence
  7.1 Existing guidance
  7.2 Evidence on the effects of interventions
  7.3 Evidence on implementation issues
8 Recommendations
  8.1 Description
  8.2 M&E and research gaps
  8.3 Dissemination
9 Review and update
  Annexes
  A.1 Search strategies for existing guidance and evidence
  A.2 Sources of existing guidance and evidence
  A.3 Management of references
  A.4 Reviews inclusion and exclusion criteria
  A.5 Appraisal of evidence: detailed evidence profiles
  A.6 Abstracts of included evidence
  A.7 Deliberative process
  A.8 Internal review
  A.9 External review
  A.10 Tools
  A.11 References
    A.11.1 General references
    A.11.2 Table of guidance retrieved
    A.11.3 Table of included evidence
    A.11.4 Table of excluded evidence

8.1.2 Dissemination strategy

In order to effectively use evidence, in this case in the form of guidance, several elements have to come together[149]:

1. Good information: soundly developed health systems guidance that can be used to address real problems.
3. Good access to information: dissemination mechanisms and systems that make it easy for end-users to access information.
4. Supportive environments: physical and intellectual environments in which research is valued and uptake of research-based knowledge is supported and encouraged.
5. Evidence-based promotion of knowledge uptake: interventions that demonstrably promote the uptake of knowledge and lead to behaviour change.

Each one of the points made above involves a wide range of possible strategies. For example, 'good information' covers a whole set of processes including generating, synthesising and translating evidence into recommendations; 'good access to information' deals with the particular needs of each audience and the most appropriate formats to effectively reach them; 'supportive environments' concern the complexity of interactions between researchers and policy makers; and 'evidence-based promotion of knowledge uptake' is grounded in theories of behaviour change.

Another source[150] describes potentially useful tips for a dissemination strategy (in this case, in the context of national quality campaigns to disseminate best practices):

- Provide simple, evidence-based recommendations
- Align messages with the strategic goals of the adopting organization
- Use a nodal organizational structure
- Engage a coalition of credible campaign sponsors
- Establish a threshold of participating organizations
- Provide practical implementation tools
- Create networks to foster learning opportunities
- Monitor progress and evaluate impact

Institutional dissemination

Institutions may already have dissemination strategies in place. Existing communication strategies, at the institutional or departmental levels, can be leverages for guidance dissemination. Institutions that produce guidance should have an institutional website where information on guidance (and guidance itself) can be found[151]. Ideally, these institutions will have an interactive repository of guidance that would require routine update and should include the whole range of materials related to guidance and guidance development, such as:

- Approaches, methods and tools.
- Guidance: produced, adapted, endorsed and in development.
- Links to other guidance related sites.
- Links to other relevant thematic sites.
- Mechanisms for user participations.
- Other typical websites components: help, news, about…

The guidance steering group may want to consider appealing to other departments or institutions and initiatives to include guidance materials into their own dissemination strategies or to make them available on their own websites. It is especially relevant that departments or units within the same institution that produces guidance are aware of new guidance and contribute to its dissemination.

For example, EVIPNet which "promotes the systematic use of health research evidence in policy-making, focusing on low and middle-income countries"[152] provides access to guidance materials[33]. It hosts a selection of policy briefs providing evidence from research syntheses in a user-friendly format on policy options. Other examples can be considered from the lists of possible sources for guidance (see Table 5 in page 20) or from the examples on evidence dissemination initiatives (Annex 12).

8.1.3 Audience and formats

The guidance 'master document' is the key output of guidance development. However, it is a sophisticated and technical document that is not appropriate for audiences unfamiliar with the approaches and methods to develop guidance. To facilitate guidance use it is paramount that other formats are used to disseminate guidance among end-users. There are studies that looked at how to present findings of systematic reviews to users (e.g. consumers[153] and other stakeholders[154]) and others that test the usability of summaries of evidence from systematic reviews[155,156].

The generic process to prepare guidance materials for end-users include (adapted from NHMRC 1999[157]):

1. Include end-users in the team adapting guidance materials.
2. Assess the need, since not all health systems guidance will be relevant for all stakeholders.
3. Define the aim according to the target audience (e.g. to inform, to form, to get them aligned with certain recommendations).
4. Draw on the expertise of a professional communicator and graphic designer and establish a budget.
5. Determine the best means (format, content, distribution strategies and implementation strategies) to reach the target audiences and the forms of evaluation that can be used.
6. Develop a list of consumers' questions that the publication should answer.
7. Review/collate the relevant materials for dissemination and prepare a first draft.
8. Assess the reading level and obtain consumer and professional reviews of the draft.
9. Revise draft to ensure all aims have been met and users' questions answered.
10. Disseminate, implement and evaluate the success of the publication.

The target audience for a dissemination and implementation strategy is the group of people who have to be influenced for change to take place[149]. Since decisions are influenced by many factors, besides the availability of evidence, the format of guidance does need to be adapted to the paradigm of each type of audience. Table 38 provides some examples of different types of audience and the formats that may be preferred for each of them.

Dissemination of guidance also has to be adapted to specific sectors of the audience, such as vulnerable groups, both in terms of contents (i.e. highlighting the actionable messages which may have a special impact on disadvantaged groups) and formats (i.e. ensuring that those affected by guidance, whether health care providers or sectors of the population, can access guidance in formats tailored to their needs)[82].

Regardless of the different formats and means of dissemination, all guidance materials should contain (with more or less detail) some basic information:

- Title of the guidance and its objective
- Outline of methods
- Recommendations
- Credentials and institutional affiliations of the guidance developers.
- Disclosure of conflicts of interest.
- Funding sources.
- Reference to the guidance 'master document'.

Table 38. Types of audience and preferred formats.

Setting	Audience	Information needs	Preferred format
Legislative	• politician • bureaucrat • interest group	• problem definition • affirmation of assumed causes • policy 'ideas'	• person-to-person • overview in brief memorandum • media
Administrative	• program manager • regional administrator • hospital executive	• program evaluation • practice variation • cost-effectiveness	• special contacts • summary report • dedicated seminar
Clinical	• practitioner • professional society • expert panel member	• effectiveness • ethics • patient preference	• colleagues • action-oriented synthesis
Consumer	• patient • friend/relative • person with chronic illness • well person	• effectiveness • safety • cost	• personal
Industrial	• company scientist • corporate executive • venture capitalist	• marketable product	• depends on scientist versus non-scientist

Adapted from Lomas 1997[158] (quoted in NHMRC 2000[125]).

The following sub-sections provide some hints on potential means to disseminate guidance[157].

a. Print

Printed materials are adequate for users who are literate and have sufficient reading comprehension. In most cases, while printing materials may be appropriate for most policy makers, they will not cover all potential users.

There are several strategies to revise complex contents into user friendly formats. The 1:3:25[159] rule, for example, consists of structuring documents to one page to convey the main messages of guidance (e.g. recommendations), three pages for an executive summary to present the body of guidance in more detail, followed by 25 pages with the full guidance.

There are several printed formats that can be used; for example, policy briefs where guidance is tailored to a local context and press releases which need to be carefully written given the brevity of the text. Brochures using drawings, colours and other graphic resources can also convey the key messages in a very attractive way (see a nice example in the reference[160]).

b. Audio-visual means

Videos can be very influential, especially if they are broadcasted through means that are accessible and popular among end-users. An example of dissemination of evidence using

video can be found in the reference[161]. However, videos of good quality are expensive to produce and, therefore, to update as well.

Audio (e.g. radio broadcasts) can also be considered, and are probably more appropriate for the general population or for health care providers.

c. Websites

Websites are very popular and provide versatile ways of storing and displaying information. They have great possibilities, such as redirecting the user to different sections by means of internal and external links and hosting nested information that could be used to successively access more complex aspects of guidance[162], such as risk of bias assessments[163]. However, they can also contain irrelevant and misleading information[164].

Websites can also host sections that allow interaction with users, such as 'blogs', comments or forums. This would be very useful to improve the dissemination of guidance materials according to users' preferences.

d. Additional considerations

Regardless of the different means of dissemination, guidance materials will have to be mindful of the contents, layout, expression and languages. These parameters are specific for each primarily targeted audience.

This type of work should be done by the experts familiar with the topic in coordination with the different experts in formatting guidance materials to ensure that the way guidance is presented does not imply changes in interpretation.

Professional editors, graphic designers, professional translators and proof-readers are involved in this phase of guidance development and their terms of reference and budget to fund this work need to be anticipated (for example, it has been estimated that one writer can process 5000 words in one week at fees of around 0.5 USD per word).

Language, even in the most technical guidance materials, has to be kept simple. Using a style guide may help in doing so from the start. Illustrations can also improve the visual aspect of the guidance materials and improve understanding of the contents.

The guidance steering group may also consider introducing case studies or stories that help explain the contents in the guidance materials or that narrate experiences of policy makers who followed the recommendations, or views of patients or health care providers.

8.1.4 Standards for dissemination

Best practices in reporting guidance have to be used, regardless of the means of dissemination. The AGREE II instrument contains items related to the reporting of guidance (Table 41). There are also other tools specific for guidance reporting[191,192]. The Conference on Guideline Standardisation (COGS) checklist for reporting clinical practice guidelines[192] (described in Table A - 10-1) is a consensus standard for reporting guidelines across organisations[165]. Most of the items are readily applicable to health systems guidance and others can be easily adapted.

In addition, specific media may have its own dissemination requirements that may be applicable for health systems guidance; such as websites on health and medical issues. For example, the 'Health On the Net Foundation" (HON) "promotes and guides the deployment of useful and reliable online health information, and its appropriate and efficient use"[166]. It offers the possibility to receive an accreditation based on the HON Code of Conduct for medical and health Web sites (HONcode)[167]. The criteria are:

- Authoritative: indicate the qualifications of the authors.
- Complementarity: information should support, not replace, the doctor-patient relationship,
- Privacy: respect the privacy and confidentiality of the personal data submitted to the site by the visitor.
- Attribution: cite the source(s) of published information, date medical and health pages.
- Justifiability: site must back up claims relating to benefits and performance.
- Transparency: accessible presentation, identifies the editor and webmaster, accurate email contact.
- Financial disclosure: identify funding sources.
- Advertising policy: clearly distinguish advertising from editorial content.

Although the focus is clinical, a similar approach could be envisaged for sites presenting guidance materials on health systems.

Research topic 5. Standards for reporting health systems guidance in several media.

Reporting standards exist for primary research and clinical practice guidelines. Most of the items contained in the latter are applicable to health systems guidance. However, health systems guidance presents several particularities, such as evidence that is primarily related to components of the health system, rather than to patients or the public in general.

- What additional items would need to be taken into account when reporting health systems guidance?
- What are the requirements of health systems guidance reporting when using websites?
- What are the requirements of health systems guidance reporting when less conventional means are used, such as video?

Research topic 6. Format and means of health systems guidance dissemination.

Effective use of evidence (health systems guidance) depends on many factors, including the access to guidance by end-users. The contents and means of delivering guidance influence the way that guidance is interpreted and used.

- What are the most efficient formats to present health system guidance to end-users to improve their uptake?
- Are those formats equally efficient across different backgrounds of end-users and institutional settings in which they operate?

8.2 Implementation

A distinction has to be made between dissemination and implementation of guidance[168] and between dissemination and implementation of research findings[169]. While policy makers may certainly need and use research findings, the assumption underlying guidance development is that guidance materials are more appropriate for policy makers as guidance is tailored to their backgrounds, has been 'digested' and is therefore more understandable, is less technical and more user-friendly and can incorporate non-research evidence, which also plays a role in the decision making processes.

Although appropriate dissemination is certainly a prerequisite for guidance implementation, it is not the only one. It is widely assumed that passive dissemination of guidance does not lead to their use[170]. Possibly, the most comprehensive systematic review on strategies for guidance dissemination and implementation[171] included 235 studies, most of them of poor quality, that looked at strategies such as: educational materials, educational meetings, consensus processes, educational outreach, opinion leaders, patient-directed interventions, audit and feedback, reminders, other professional interventions (including mass media and marketing), financial interventions, organisational interventions, structural interventions and regulatory interventions, both alone or multifaceted. The majority of studies focused on clinical guidelines and showed improvements in care, although with great variability between and within interventions. A most recent review on the effects of interventions for seeking, appraising and applying evidence from systematic reviews in decision making by policymakers and managers found only weak evidence that these interventions had positive effects[172].

A search in the Cochrane Library* identified a few systematic reviews potentially relevant to guidance implementation. One review found that guidelines can improve care provided by professionals allied to medicine[173]. A review[174] about interventions to improve outpatient referrals found that passive dissemination of local referral guidelines was ineffective but that dissemination of guidelines with structured referral sheets was effective. Another review[175] found that tailored implementation strategies† are more likely to improve professional practice than no strategy or dissemination of guidelines alone.

A review on the effects of audit and feed-back on professional practice reported a multi-centre study which found little effect of audit and feed-back with educational meetings (compared to no intervention) at improving compliance with guidelines for asthma. Two other studies comparing audit and feed-back alone with audit and feed-back plus incentives and a practice-based seminar, respectively, found no differences in guidelines implementation. Another study compared audit and feed-back with reminders or with local opinions leaders and both found that comparison interventions were more effective in complying with guidelines. Another study compared audit and feed-back with practice-based education, showing no differences in complying with guidelines[176].

A review looking at interventions to improve professional adherence to guidelines for prevention of device-related infections is in a protocol stage[177].

There are several study designs to evaluate guidance implementation strategies[178]. However, in the case of clinical guidance, patients, practices or health facilities can be randomised or selected to test adherence to guidance by health care providers. In the case of health systems guidance, this can be much more difficult since the scope of guidance is the health systems itself (or components of the health system) that can hardly be 'atomised' in geographical or other terms to allow random selection.

---

* "guid* AND (imple* OR dissem*)" in Title, Abstract or Keyboards (April 2011).

† Strategies that are planned taking account of prospectively identified barriers to change.

Issues that may be taken into account when thinking of guidance implementation include:

- analysis of local needs and priorities;
- barriers and facilitators to implementation;
- resources for implementation;
- strategy to support the adoption of recommendations.

Certain components of guidance could certainly facilitate implementation and appraise the 'implementability' of guidance[179,180]. However, guidance implementation is rather an organisational endeavour[181], i.e. guidance implementation strategies are not really specific to the particular topics for which guidance is developed. Implementation strategies can be linked to other organisational improvements, such as quality assurance initiatives or programmes for change[182]. A recent review[150], for example, identified eight strategies for the dissemination of evidence-based practices in the context of national quality campaigns*.

Finally, guidance implementation should also be monitored[183] to ascertain to which extent recommendations have been followed and to study the factors that may have facilitated or obstructed implementation. Again, this type of activity would require an organisational effort rather than a topic specific approach.

Research topic 7. Strategies to promote the uptake of health systems guidance by policy makers.

While there is some research on strategies to implement clinical guidelines and to promote adherence to them, there is less robust research on how to implement health systems guidance among policy makers. Experimental or quasi-experimental designs may not be feasible.

- What is the 'menu' of (organisational) strategies to promote adherence to health systems guidance by policy makers?
- What strategies are most effective in promoting uptake of health systems guidance by policy makers?
- What study designs can provide robust evidence on the effects of interventions to promote the uptake of health systems guidance among policy makers?

* (1) Highlight evidence base and relative simplicity of recommended practices. (2) Align the campaign with the strategic goals of the adopting organizations. (3) Increase recruitment by integrating opinion leaders into the enrolment process and employing a nodal organizational structure. (4) Form a coalition of credible campaign sponsors. (5) Generate a threshold of participating organizations that maximizes network exchanges. (6) Develop practical implementation tools and guides for key stakeholder groups. (7) Create networks to foster learning opportunities. (8) Incorporate monitoring and evaluation of milestones and goals.

9 How to plan for the update of guidance and how (and what) to review guidance

This is our fate: to stand
In our own way. Forever
in the way.
(Rainer Maria Rilke, The Eighth Elegy)

The final stages in guidance development are to plan for its update and to undertake external and internal reviews of guidance. The update plan can only be carried out at the end, when the previous work has been completed and all aspects that might need to be revisited have been made explicit. Finally, guidance is submitted to external and internal review processes to identify errors, inaccuracies or gaps and to improve its quality.

9.1 Update

Update of guidance is justified by the fact that new evidence can be made available at any time to fill an evidence gap or to reinforce or to contradict existing evidence. Furthermore, new research in the methods used along the whole guidance development process may well be refined over time (e.g. to assess the quality of evidence, to combine quantitative and qualitative evidence); or advances in the understanding of health systems and how interventions operate across all components of the system may occur as well.

Criteria for deciding when to update guidance may include:

- the date of the most recent evidence: this refers to the date of the systematic review that provided the evidence on the effects of interventions (as opposed to the reviews, if different, that provided evidence on implementation issues). The publication date of the review, the range of years covered by the search strategy of the primary research included in the review and the date of the most recent primary research should all be looked at. A systematic review could have been published in 2010 but the search strategy of the primary research may have reached only up to 2008, and the most recent study included in the review dated from 2006. This guidance would be based on five year old evidence;
- indications that new evidence may be available shortly; for example, the existence of a protocol of a relevant Cochrane review, published in the Cochrane Library; or based on knowledge of ongoing research projects;
- when developing guidance using rapid methods*, in this case updating may be especially relevant
- a reasonable timeframe that allows to incorporate recent research findings or lessons learned from similar guidance implementation.

Secondly, it should be considered that not all aspects of guidance may need to be updated at the same time. For example, most typically, evidence profiles (and recommendations based on them) will need to be updated based on new evidence. However, if new methods are available, only certain parts of the guidance will need to be revisited.

It is advisable to outline updating issues into a guidance update plan and to state in the guidance when minor or major updates are expected to take place (e.g. by a statement such as “review by [date]”). An outline of the update plan should be part of the final guidance document.

* These are not covered in this handbook. See Research topic 2 (page 25).

9.2 Reviewing guidance

The final stage in guidance development before actually publishing and disseminating guidance materials is to review it. The objective of reviewing guidance is to improve the quality of guidance by systematically scrutinising the methods, processes and outputs of guidance development.

There can be external and internal reviews. External reviews provide a 'fresh' feed-back on the processes and outputs of guidance, while the internal review can also reveal procedural issues in guidance development that may need to be improved. External and internal reviews should refer to accepted guidance standards provided they are available. Similar to the assessment of quality of systematic reviews, both the quality of reporting and the quality of the recommendations can be assessed.

9.2.1 External review

The WHO handbook for guideline development[1] establishes several critical steps in the guidance development process for peer-review (and who could review them), namely:

- questions that triggered guidance development: experts and end-users at WHO headquarters, regional offices and externally;
- protocol of systematic reviews commissioned: circulated to experts for comments on the methods and evidence identified;
- evidence profiles and tables: experts for identification of any missing evidence;
- documentation used by panellists for developing recommendations: experts and organizations representatives of relevant stakeholders;
- guidance draft with recommendations before publication.

a. Reviewing guidance

Health systems guidance is complex in terms of the sequence of processes involved in its development (from defining a problem to issuing recommendations), the wide range of approaches applied (e.g. assessing the quality of evidence, deliberative processes), the methods used (i.e. related to quantitative and qualitative evidence) and the plethora of additional issues to take into account (e.g. evaluation of recommendations or implementation of guidance). Therefore, reviewing guidance could pose major challenges to any reviewer, if nothing else due to the amount of time needed to complete a full review. Once the guidance is completed, the guidance steering group should decide whether to have reviewers for the full guidance or reviewers for specific components of the guidance. Other issues to decide are the number of reviewers and their profiles.

The table below serves as suggestions on how the external review could be articulated:

Table 39. Suggestions for external reviews of guidance.

Steps	Review objective	Reviewers profile	Approach
Rationale for guidance	To agree on the need to develop guidance	• Topic experts • Decision makers • Users associations	Check the wording of the problem and its framing in the wider health system framework.
Guidance protocol	To ensure that guidance is appropriately addressing the health system problem	• Researchers, methodologists • Topic experts • Decision makers • Users of health services	Split the protocol into the relevant parts for each type of reviewer.
Evidence profiles (effects of interventions and implementation issues)	To verify that all relevant evidence has been retrieved and properly assessed and interpreted	• Researchers, methodologists • Topic experts • Decision makers • Users of health services	Emphasis on the interpretation and implications of the evidence profiles.
Recommendations	To ensure that recommendations match existing evidence	• Topic experts • Decision makers • Users of health services	Verify the explicit links between each recommendation and the evidence supporting it.
Full guidance	All of the above	• Topic experts • Decision makers • Users associations	AGREE II instrument (see 9.2.2)

As seen in Table 39, if there are only two reviewers in each category, the number of reviewers would be between six and eight depending on the step of the review, which is a considerable number in terms of collating (potentially contradictory) feed-back. In order to minimise the burden of the external review without compromising its usefulness, reviewers should be provided with clear guides (e.g. questionnaires) and in some cases draw from groups of people or institutions rather than individuals (e.g. users or professional associations). Conflicts of interests have to be cleared as well (see section 10.2).

The review process should be transparent and objective. Peer-review is not a straightforward activity and needs a systematic approach[184] supported by review guides. Review guides should assist the review process and lead to useful suggestions on ways to improve the guidance being reviewed. The process of peer-reviewing research articles for bio-medical journals can provide some hints on guides for reviewers (for example, BioMed Central[185] and the British Medical Journal[186]).

The review guide should be tailored to each critical step in the guidance development process that needs to be reviewed. When full guidance is reviewed the AGREE II instrument can be used at the end (see 9.2.2).

Guides to reviews should include (i) a generic and brief background of the guidance project; (ii) the objective of the review, what is expected from the review; (iii) clear instructions on how to undertake the review, including the format of the expected feed-back; (iv) the document or part of the document to be reviewed; (v) the timeline; and (vi) the contact person in the guidance steering group.

Some examples of questions to guide the external review process in each step include:

- Rationale for guidance
    - Is the problem statement clear? If not, suggest an alternative statement.
    - Is the problem well circumscribed within a health system component? If not, suggest a way of relating the problem to a health systems component.
    - Is the problem properly related to an appropriate health system framework? If not, suggest links of the problem to a health system framework.
    - How important is the problem? Please provide a justification for your opinion.
    - Are you aware of any existing guidance which has already addressed the problem? If yes, please provide the full reference and a link to the complete version if available.
    - If not, do you think that it is worthwhile to develop guidance? Please provide a justification for your opinion.
- Guidance protocol: use or adapt an existing checklist to assess protocols; for example, the EPOC Checklist for Refereeing Protocols for Reviews[187] (adapted in the following list):
    1. Is the specific purpose (question to be answered) stated?
    2. Are the participants (providers and users) clearly stated?
    3. Are the sources and search methods used to find evidence on the questions to be answered stated?
    4. Is the search strategy for evidence reasonably comprehensive?
    5. Are explicit criteria used for deciding which evidence to include?
    6. Is bias in the selection of evidence likely to be avoided?
    7. Are there reasons for excluding evidence?
    8. Are the criteria used for assessing the quality of reporting and the methodological quality of evidence reported?
    9. Is the quality of all evidence assessed using appropriate criteria?
    10. Are the methods used (regardless of whether they are qualitative or quantitative) to combine the evidence and to issue recommendations reported?
    11. Are the methods used to combine the evidence and to issue recommendations appropriate to the questions to be answered by the guidance?
- Evidence profiles
    - Are evidence profiles consistent with the terms of the scope of guidance?
    - Are evidence profiles consistent with the underlying sources of evidence?
    - Are evidence profiles explicit about what evidence exists and about evidence gaps as well?
    - Is the quality of evidence made explicit?
    - Are the interpretations of evidence profiles appropriate?
    - Are evidence profiles understandable, clear and user-friendly (for users, included)?
    - Are the methods to develop evidence profiles described and followed?

- Recommendations
    - Are recommendations consistent with the evidence available?
    - Is the strength of recommendations qualified using an accepted scale?
    - Do recommendations incorporate issues related to implementation capacity in general, including feasibility and sustainability?
    - Are recommendations complemented with indications on how to monitor and evaluate their implementation?
    - Are the methods to develop recommendations described and followed?

b. Incorporating feed-back from the reviews

The review process should be documented by keeping track of the very documents that were sent out for review, the responses received and the changes that the guidance documents have undergone.

The guidance steering group should collate all feed-back received at each critical step of the guidance development process and engage with reviewers where inconsistencies arise. The group should prepare a list of suggestions from reviewers and propose which of those suggestions should be taken on board and which ones should be disregarded. Decisions should be submitted to the guidance development group and changes approved. If needed, a second round of reviews could be justified if an especially controversial issue is raised.

Versions of the guidance materials being developed should be kept separate for transparency and documenting purposes, with a clear coding system to designate different versions of the documents.

9.2.2 Internal review

The guidance steering group may wish to undertake an internal final review as a wrap up of the whole process of guidance development. The objectives of this internal review are:

1. to further improve the quality of guidance, in terms of its contents;
2. to further improve the quality of reporting of guidance;
3. to draw lessons learned from key processes of guidance development in order to inform future improvements on the efficiency of the processes.

Several tools can assist the internal review:

- Formal and informal feed-back from participants in the guidance development processes (feed-back on the guidance development process).
- Systematic review of the quality of guidance using an assessment tool (see Table 41, page 116).

a. Feed-back from participants in the guidance development processes

The guidance steering group may consider several parameters to be assessed by the participants in the guidance development processes, such as:

- General management of the processes
    - Leadership (applicable to each of the several groups of people or experts)
        - Capacity to lead the group towards a common endeavour
        - Coordination between the different teams or groups, including adequate information flow
        - Communication strategy with stakeholders not involved in the guidance development process but with potential interests in it
        - Efficient fund raising
    - Resources
        - Optimal timing between the different phases of guidance development
        - Efficient use of time
        - Availability of office space and resources to undertake the tasks
- Activities (including meetings, teleconferences)
    - Preparation
        - Clarity of objectives or targets to be achieved
        - Degree of success in achieving the objectives of the meeting
        - Agenda
    - Execution
        - Appropriateness of approaches
        - Availability, clarity and timeliness of documentation
        - Information and communication technologies
        - Availability of resources
        - Staff performance
    - Results
        - Degree of achievement of objectives or targets
        - Participants' satisfaction
        - Relevance of the results
- Documentation (especially the guidance development handbook)
    - Presentation
        - Clarity of the statements and language
        - Appropriate distinction between what is essential and what is accessory
        - Adequate length, overall and by chapters
        - User friendliness
    - Contents
        - Completeness in the descriptions of concepts, approaches, methods and outputs
        - Adequate balance between explanations in the handbook and additional information to be sought elsewhere
        - Up to date and relevant references
    - Overall
        - Directness between the reading of the handbook and the capacity to undertake the activities outlined in it
        - Overall satisfaction with the handbook as a tool to develop guidance

This feed-back can be obtained by either formal or informal means. Care has to be taken not to overload participants with additional tasks, such as filling lengthy assessment questionnaires. A mix of formal and informal methods can be used. The guidance steering group can be especially attentive to informal feed-back during meetings and other exchanges, and may want to record them. During key events (e.g. in the deliberative process to issue recommendations) it could be useful to gather formal feed-back from participants.

Table 40 provides suggestions on approaches for informal and formal review for each group of participants in the guidance development process. The items and approaches used in the internal review depend on the potential use of them; ultimately, on the way that guidance production is institutionalised (e.g. as a 'team', 'departmental' or 'institutional' activity).

Table 40. Approaches to undertake an internal review of guidance development processes.

	Main functions	Management review	Activities review	Documentation review
Guidance steering group with coordinator	Manage guidance development processes (coordinator) with the support of the members	Formal review of management. Especially important since the guidance steering group 'manages' the whole process	Formal review	Formal review of the handbook and other materials
Guidance development group	Assists and endorses scoping of guidance, evidence profiles and recommendations	Formal review of management	Informal review of activities, since the group is not systematically involved in the routine of guidance development process	Formal review of key documents (e.g. evidence profiles); informal review of relevant parts of the handbook
Guidance panel for recommendations	Issuing of recommendations	Formal review of management, especially the facilitation aspect of the deliberative process	Formal review of the deliberative proves	Formal review of documents used to work on recommendations (e.g. evidence profiles)
External reviewers	Reviews scope of guidance, evidence profiles and recommendations	Not applicable	Not applicable	Formal review of the guidance sub-products, since this is their mission

	Main functions	Management review	Activities review	Documentation review
Technical support	Specific technical tasks	Informal review of management, specifically in terms of liaising specific technical tasks with the whole guidance development process	Informal review of specific tasks	Informal review of input documentation to undertake the technical support tasks

b. Systematic review of guidance using an assessment tool

There are several tools to assess the quality of guidance[188,189,190], some of them focusing on the particular aspect of reporting[191,192]. The AGREE II instrument[190,193] has shown the capacity to discriminate higher quality guidelines that follow technical documentation from those of lower quality[11] and it is listed among the online resources for clinical guidelines in the Cochrane Collaboration site[194]. It is suggested here to use the AGREE II instrument as the internal review tool. The purposes of AGREE II are to[195]:

1. assess the quality of guidelines;
2. provide a methodological strategy for the development of guidelines; and
3. inform what information and how the information ought to be reported in guidelines.

AGREE II is designed for clinical and public health guidelines, including health promotion, public health, screening, diagnosis, treatment or interventions. It has not been designed to assess the quality of guidance of 'health care organisational issues'[195]. Table 41 shows the domains and items of the AGREE II instrument[195] with suggestions on how some items could be adapted to health systems guidance.

The AGREE II user manual contains instructions on how to apply the tool and links to references. It consists of 23 items organised within six domains and an overall assessment:

- Domain 1. Scope and Purpose is concerned with the overall aim of the guideline, the specific health questions, and the target population (items 1-3).
- Domain 2. Stakeholder Involvement focuses on the extent to which the guideline was developed by the appropriate stakeholders and represents the views of its intended users (items 4-6).
- Domain 3. Rigour of Development relates to the process used to gather and synthesize the evidence, the methods to formulate the recommendations, and to update them (items 7-14).
- Domain 4. Clarity of Presentation deals with the language, structure, and format of the guideline (items 15-17).
- Domain 5. Applicability pertains to the likely barriers and facilitators to implementation, strategies to improve uptake, and resource implications of applying the guideline (items 18-21).
- Domain 6. Editorial Independence is concerned with the formulation of recommendations not being unduly biased with competing interests (items 22-23).

- Overall assessment includes the rating of the overall quality of the guideline and whether the guideline would be recommended for use in practice.

AGREE II assessments should be undertaken by two (ideally four) assessors. Each item is given a score between 1 and 7 (from 'strongly disagree' to 'strongly agree'). The user manual specifies the criteria to assess each item. For each domain an averaged score is calculated as a percentage.

As suggested in Table 41, the domain 'Scope and purpose' may need adaptation for health systems guidance because this is precisely what defines certain guidance as 'health systems' guidance. While scoping clinical guidelines has to do with diagnosing or treating patients affected by a specific condition, health systems guidance addresses issues related to the organisation of the services and how systems are set up to deliver health care. Stakeholders and users are also different: they are mainly patients and clinicians for clinical guidelines, and health care providers (and also users) and decision makers and managers for health systems guidance.

The other domains are methodological or related to implementation and are equally relevant to health systems guidance: 'Rigor of development', 'Clarity of presentation', 'Applicability' and 'Editorial independence'. The 'different options for management of the condition or health issue are clearly presented' item (domain 4) should include, in the case of health system guidance, considerations on the different health system-level interventions that focus on specific health systems components (as framed in the problem definition addressed by guidance).

Research topic 8. Adaptation of AGREE to health systems guidance.

While AGREE II has shown capacity to discriminate between high and poor quality guidelines[11], the instrument has a clear focus on clinical guidelines. The performance of AGREE II in health systems guidance may be tested against health systems guidance.

Table 41. The AGREE II instrument: application to health systems guidance.

Domains and items	Explanation	Adaptation to guidance on health systems	Chapter or section in the handbook
Domain 1. Scope and purpose			
1. The overall objective(s) of the guideline is (are) specifically described.	Potential health impact of guidance; problem addressed.	Although it is expected and assumed that health systems interventions will have an impact on the health status of the population, health systems guidance targets health systems components.	2.1 Framing the 'problem' for health systems guidance
2. The health question(s) covered by the guideline is (are) specifically described.	Health questions particularly for the key recommendations.	Detail of the health system problem and the possible health system-level interventions.	
3. The population to whom the guideline is meant to apply is specifically described.	Demographic characteristics of the population.	(a) Health systems components targeted by the intervention; (b) population that is meant to ultimately benefit from health systems changes.	
Domain 2. Stakeholder involvement			
4. The guideline development group includes individuals from all relevant professional groups.	Including members of the steering group, research team, panellists.	No need for adaptation.	10.2 People
5. The views and preferences of the target population (patients, public, etc.) have been sought.	Target population experiences and expectations.	Consider views and preferences of (a) health care providers, regulators, policy makers and (b) population.	6 How to translate evidence into guidance and recommendations
6. The target users of the guideline are clearly defined.	Users of guidance, those who will implement it; clinicians.	Policy makers and managers are the primary audience of health systems guidance.	3.1.2 Primary and secondary audiences (decision level)
Domain 3. Rigour in development			
7. Systematic methods were used to search for evidence.	Sources and search strategy.	No need for adaptation.	4.1.1 Searching for evidence
8. The criteria for selecting the evidence are clearly described.	Explicit inclusion and exclusion criteria.	No need for adaptation.	4.1.2 Relevance and inclusion of evidence

Domains and items	Explanation	Adaptation to guidance on health systems	Chapter or section in the handbook
9. The strengths and limitations of the body of evidence are clearly described.	Quality of evidence, risk of bias.	No need for adaptation.	5 How to assess and present evidence
10. The methods for formulating the recommendations are clearly described.	Such as voting, formal and informal consensus techniques.	No need for adaptation.	6 How to translate evidence into guidance and recommendations
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	Evidence that these issues have been addressed.	No need for adaptation.	
12. There is an explicit link between the recommendations and the supporting evidence.	Relation between evidence and each specific recommendation.	No need for adaptation.	
13. The guideline has been externally reviewed by experts prior to its publication.	External reviewers not involved in guidance development.	No need for adaptation.	9.2.1 External review
14. A procedure for updating the guideline is provided.	Explicit mention of the approach for updating.	No need for adaptation.	9.1 Update
Domain 4. Clarity of presentation			
15. The recommendations are specific and unambiguous.	Which option is appropriate. Make explicit when there is uncertainty.	No need for adaptation.	6.3 Outputs: recommendations
16. The different options for management of the condition or health issue are clearly presented.	Consider screening, prevention, diagnosis, or treatment.	Consider health system components and entry points for interventions recommended.	
17. Key recommendations are easily identifiable.	Responsive to the main problem and clearly presented.	No need for adoption.	
Domain 5. Applicability			
18. The guideline describes facilitators and barriers to its application.	Examples provided are related to health systems barriers.	No need for adaptation.	6.3 Outputs: recommendations
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	Using guidance and additional material.	No need for adaptation.	8.1 Dissemination
20. The potential resource implications of applying the recommendations have been considered.	Examples provided are related to health systems barriers.	No need for adaptation.	6 How to translate evidence into guidance and recommendations

Domains and items	Explanation	Adaptation to guidance on health systems	Chapter or section in the handbook
21. The guideline presents monitoring and/or auditing criteria.	Measuring the application of guideline recommendations.	No need for adaptation.	8.2. Implementation
Domain 6. Editorial independence			
22. The views of the funding body have not influenced the content of the guideline.	Explicit statement.	No need for adaptation.	10.2. People
23. Competing interests of guideline development group members have been recorded and addressed.	Explicit statement.	No need for adaptation.	

10 How to manage the guidance development processes

Management is getting things done.
(McMahon R, Barton E, Piot M. On being in charge.
A guide to management in primary health care. WHO Geneva 1992)

10.1 Institutionalisation

A guideline programme is defined as "a structured and coordinated programme designed with the specific aim of producing several [health systems] guidelines"[196]. As a programme it should ideally be institutionalised, given that guidance development involves a substantial amount of work over a prolonged period of time and a significant number of people. The guidance programme, with its own budget, should be supported by management, infrastructure and staff of a wider organisation.

WHO has a Guidance Review Committee established in the year 2007 by the Director General with the mission of ensuring transparency in guidance development methods and compliance with evidence-based decision making processes[197]. Several organisational models can be conceived hosting a mix of functions similar to those of the GRC. The scope of activities that could be undertaken by the GRC is certainly endless, including the managerial support of guidance development (see the other sections in this chapter for examples).

10.2 People

Table 42 describe the groups of people involved in guidance development, their membership and functions. The guidance steering group, with a coordinator, manages the whole process and ensures that the guidance development plan is implemented.

External expertise will be needed at several stages, such as literature search experts and translators. Depending on the composition of the guidance steering group, its members may be able to assume some of the more technical tasks; otherwise they will have to be externalised. The need for external inputs will also depend on the type of support that the GRC can provide.

It is a WHO requirement that no individual can be involved in guidance development prior to having been cleared for competing interests[1] (see specific WHO regulations for further details in complying with this requirement).

All personnel involved in guidance development should work under clear terms of reference that should be signed upfront and include remunerations such as per diems, travel and accommodation, where applicable. The groups outlined in Table 42 should also have clear parameters such as the number of members, their constituencies, qualifications, membership requirements and selection and substitution mechanisms.

Table A - 10-2 (in Annex 7) provides a form to assist in the management of teams and groups involved in guidance development. The table contains columns to register the availability of curriculum vitae, signed terms of reference and declaration of interest forms.

Table 42. People involved in guidance development and their functions.

	Members	Functions
Guidance steering group with coordinator	• Staff from the institution developing guidance, including the main departments with an interest on the scope of guidance	• Manage the whole guidance development processes (coordinator) with the support of the members • Ensures adherence to best practices (coordinator) • Undertakes technical tasks
Guidance development group	• Expert(s) in health systems • Expert(s) in the topic of the guidance • Expert(s) in research synthesis and in knowledge translation • Representatives of stakeholders • Representative of potential users	• Assists in the scoping of guidance, and endorses it • Reviews the completeness of available evidence • Reviews and endorses the assessments of the quality of evidence • Reviews and endorses the recommendations • Supervises adherence to best practices
Guidance panel for recommendations	(same constituencies as for the members of the guidance development group but willing and able to participate in the deliberative process)	• To issue recommendations based on the available evidence (panel facilitator, coordinator, members of the development team, of the steering group and technical support)
External reviewers	(same constituencies as for the members of the guidance development group but as external resources to provide 'fresh' and independent advice)	• Advice on the scope of guidance • Provide sources of guidance and of evidence • Verify adherence to best practices • Verify relevance of the guidance • Suggest ways to improve guidance
Technical support	• Literature searches specialist • Researchers, statisticians • Communication experts and workshop facilitator • Editors, graphic designers, proof-readers • Translators • Other (can be external)	• To provide ad hoc external technical assistance as needed.

10.3 Knowledge management

The whole process of guidance development involves the management of eventually large amounts of information and documents. It is advisable to have explicit documentation management rules and possibly a knowledge management strategy embedded in the institution. Knowledge management could be supported by specific tools designed for guidance development, such as the 'Guideline Elements Model'[198], or systems specifically developed for this purpose by the institution.

The documentation of the guidance development process can be classified into four categories:

- Management documentation
- Technical documentation
- Guidance master document
- Guidance protocol

The guidance protocol (Table 8, page 34) and master document (Table 37, page 99) have already been detailed. The management and technical documentation are listed in the tables below.

Table 43. Table of contents of the management documentation.

1	Gantt chart	
2	Resources	
	2.1	People
		2.1.2 People checklist
		2.1.3 People individual files
		2.1.4 Terms of reference
		2.1.5 Groups guides
		2.1.6 Personal documentation
		2.1.7 Declaration of interest
	2.2	Finances
		2.2.1 Financial checklist
		2.2.2 Expenses
		2.2.3 Budget
		2.2.4 Financial documents
3	Tools	
	3.1	Checklist for the rationale of guidance
	3.2	Quality of guidance (AGREE II)
	3.3	Quality of reviews (AMSTAR)
	3.4	Quality of reporting of reviews (PRISMA)
	3.5	Forms and templates
4	Clearance documents	

Table 44. Table of contents of the technical documentation.

1 Guidance handbook
2 Problem definition
2.1 Data / information for problem definition
3 Scope of guidance
3.1 Data / information for scoping
3.2 Guidance search strategies
3.3 Output files from searches
3.4 References management database
3.5 Inclusion / exclusion checklist
3.6 Quality assessments
4 Evidence retrieval
4.1 Search strategies
4.2 Output files from searches
4.3 References management database
4.4 Inclusion / exclusion checklist
4.5 Quality assessments
5 Evidence profiles
6 Deliberative process
6.1 Guidance for panellists
6.2 Minutes of meetings and other exchanges
7 Monitoring and evaluation approaches
7.1 Data / information to support M&E approaches
8 Dissemination plan
8.1 Raw materials
8.2 Documentation of designs and media
9 Reviews of guidance
9.1 Internal - AGREE II documentation
9.2 External
9.2.1 Documents sent
9.2.2 Feed-back
9.2.3 Tracked changes
9.3 Update plan
Annexes
A.1.1 Included guidance
A.1.1 Excluded guidance
A.1.2 Included evidence
A.1.3 Excluded evidence

10.4 Generic plan and timeline

Figure 19 shows a generic Gannt chart that can be used to plan and follow up on the guidance development processes. It also contains the main components of each individual process in terms of inputs, process and outputs.

Figure 19. Gannt chart of guidance development.

Tasks	Inputs	Process / Tools	Outputs	Months 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1 Familiarisation with guidance development	Handbook	Handbook	-	
2 Rationale for health systems guidance				
Framing the problem				
Modelling and defining parameters	Initial scope	Frameworks, indicators	Problem framed	
Need for guidance				
Deciding on the need for guidance	Problem framed	Screening checklist	Decision to proceed	
Searching for guidance	Initial scope	Search strategy and sources	Guidance retrieved	
Appraising guidance	Guidance retrieved	AGREE II	Guidance available	
Deciding on the availability of guidance	Guidance available	Screening checklist	Decision to proceed	
Initial approval				
3 Scoping health systems guidance				
Scoping guidance	Problem framed	PICO questions	Scope of guidance	
Writing the protocol	Scope of guidance	Follow scope parameters	Guidance protocol	
Protocol clearance				

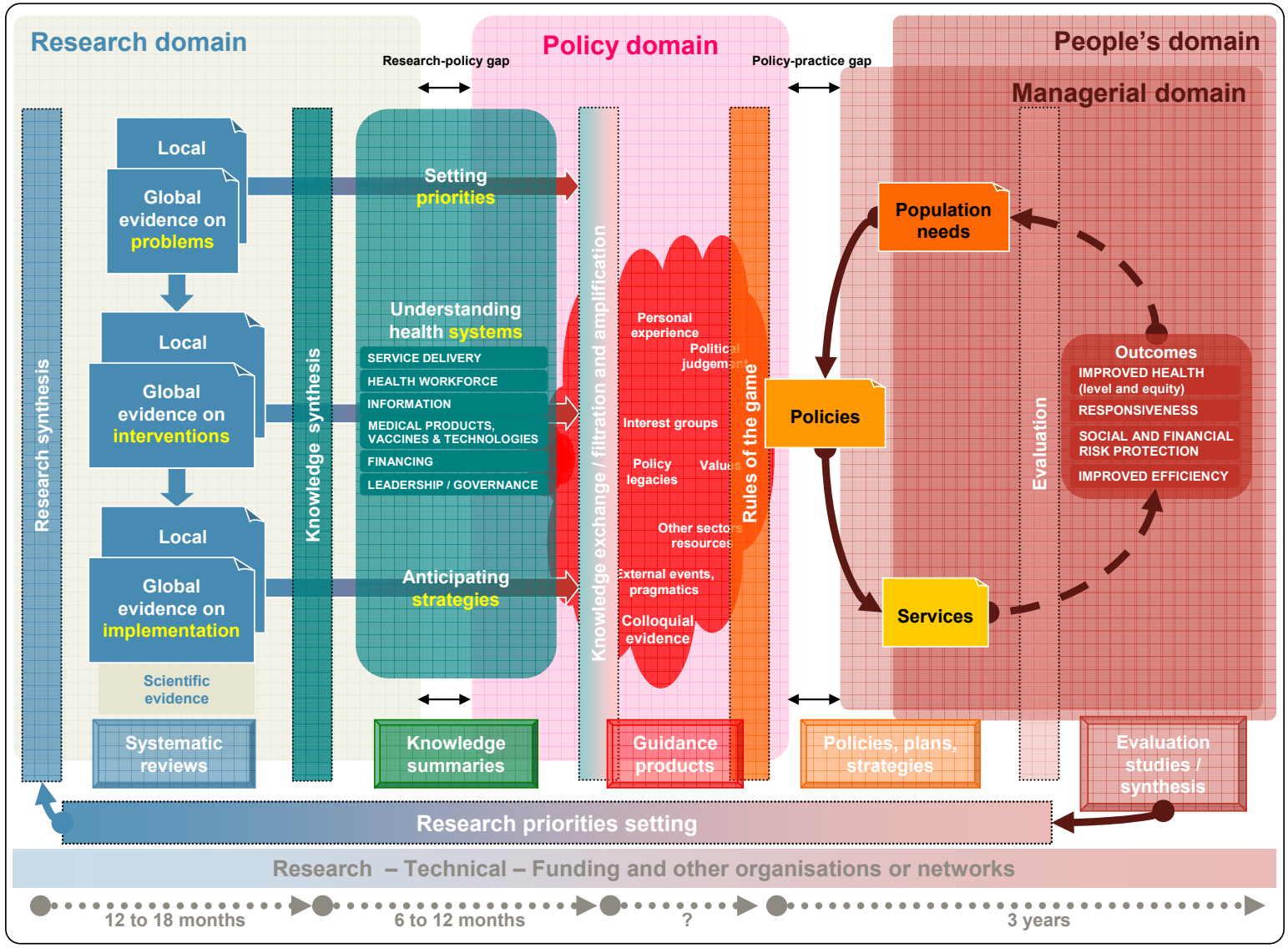
Tasks	Inputs	Process / Tools	Outputs	Months 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
4 Search and retrieval of evidence				
on the effects of interventions	Search strategy	AMSTAR, PRISMA and other	Evidence synthesis	
on implementation issues	Search strategy	AMSTAR, PRISMA and other	Evidence synthesis	
5 Assessment of evidence				
on the effects of interventions				
Appraise evidence	Evidence gathered	GRADE	Synthesis of evidence	
Present evidence	Synthesis of evidence	Communication	Evidence profile	
on implementation issues				
Assess issues	Evidence gathered	ExpandNet	Synthesis of issues	
Present issues	Synthesis of issues	Communication	Intervention profile	
Clearance of evidence profiles				
6 Translation of evidence into recommendations				
Combining evidence	Evidence profiles	GRADE, permutation matrix...	Evidence combined	
Deliberative process	Evidence profiles and combined	Consensus methods	Recommen-dations	

	Tasks	Inputs	Process / Tools	Outputs	Months 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
7	**Monitoring and evaluation of recommendations**				
	Setting up M&E modalities	Recommen-dations	M&E approaches	M&E recommen-datios	
8	**Dissemination and implementation of guidance**				
	Pre-dissemination stage	Guidance	Report writing	Master document	
	Dissemination	Master document	Didactics, design...	Guidance materials	
9	**Review and update guidance**				
	Update plan	Master document	Analysis of evidence	Update plan	
	External review	Master document	Check-lists	Feed-back	
	Internal review	Master document	AGREE II	Final master document	
	Final approval				
10	**Management**				
	Closing the guidance development process	Management documents	Administration	-	

ANNEXES

Annex 1. Framework for health systems guidance

Figure A - 1. Analytical framework for health systems guidance.



Annex 2. Glossary

This is a glossary of terms that are commonly used in the fields of evidence based medicine, research synthesis, guidance and epidemiology. Adapted from SUPPORT Tools for evidence-informed health policy making[199] (the article refers to other glossaries in these fields as well).

Acceptability	The extent to which the indicator is acceptable to those being assessed and those undertaking the assessment.
AMSTAR	A tool designed to assess the quality of the methods used to conduct a systematic review.
Baseline conditions / characteristics	Values of demographic, clinical and other variables collected for each participant at the beginning of a trial, before the intervention is administered.
Baseline risk	The risk (likelihood) of an outcome without implementing an intervention or at the beginning of a study.
Burden of disease	This refers to the impacts (or burden) of a health problem or condition (such as hypertension) in an area (such as a country or province), as measured by mortality, morbidity or other indicators. Burden of disease is sometimes measured using the 'disability-adjusted life year' or DALY – a time-based measure combining years of life lost due to premature mortality and years of life lost due to time lived in states of less than full health.
Case survey	A method for synthesising findings from a number of qualitative studies or to combine qualitative and quantitative evidence within a single review. It involves the systematic coding of relevant data from the included qualitative case studies and the subsequent conversion of these codes into a quantitative form. This then allow statistical analysis.
Confidence interval	A confidence interval is a range around an estimate that conveys how precise the estimate is; for example an estimate of the risk of an event occurring or an estimate such as a risk ratio that compares the risk with and without an intervention. The confidence interval is a guide to how sure we can be about the quantity we are interested in. The narrower the range between the two numbers, the more confident we can be about what the true value is; the wider the range, the less sure we can be. The width of the confidence interval reflects the extent to which chance may be responsible for the observed estimate (with a wider interval reflecting more chance).
Control group	A group of participants in a study not receiving a particular intervention, used as a comparator to evaluate the effects of the intervention (see 'Intervention group').
Controlled before-after study	A non-randomised study design where a control population of similar characteristics and performance as the intervention group is identified. Data are collected before and after the intervention in both the control and intervention groups.
Controlled trial	(see trial)
Cost-effectiveness analysis	An economic evaluation in which the costs and consequences of alternative interventions are expressed cost per unit of health outcome (e.g. cost per additional stroke prevented).

Cross-case techniques	A method for synthesising findings across a number of qualitative studies or cases.
Disaggregated data	Often data are reported for whole populations or areas. This can be referred to as aggregate data. In some cases, it may be desirable and possible to further breakdown or analyse these data to look at specific groups (such as people over the age of 65 years) or areas (such as a health district). This can be referred to as disaggregated data.
Effectiveness	The extent to which a specific intervention, when used under ordinary circumstances, does what it is intended to do.
Empirical evidence	Empirical results based on observation rather than on reasoning alone.
Equity considerations	Attention to how a policy or programme may impact on inequities (see inequity).
Estimate of effect	The observed relationship between an intervention and an outcome expressed as, for example, a number needed to treat to benefit, odds ratio, risk difference, risk ratio, standardised mean difference, or weighted mean difference.
Evaluation	a term often used interchangeably with monitoring. The former usually suggests a stronger focus on the achievement of results.
Evidence	An appearance from which inferences may be drawn; an indication, mark, sign, token, trace. Ground for belief; testimony or facts tending to prove or disprove any conclusion (OED[100]).
Evidence-based medicine (EBM)	Evidence-based medicine is the conscientious use of current best evidence in making decisions about the care of individual patients or the delivery of health services. The terms 'evidence-based health care' and 'evidence-based practice' are often used interchangeably with 'evidence-based medicine'.
Evidence-informed health policymaking	Evidence-informed health policymaking is an approach to policy decisions that aims to ensure that decision making is well-informed by the best available research evidence. It is characterised by the systematic and transparent access to, and appraisal of, evidence as an input into the policymaking process.
Feasibility	The extent to which valid, reliable and consistent data are available for collection.
Funnel plot	A graphical display of some measure of study precision plotted against effect size that can be used to investigate whether there is a link between study size and treatment effect. One possible cause of an observed association is reporting bias. The plot is therefore often used to assess whether publication bias is likely within a systematic review.
Grey literature	Grey literature is the kind of material that is not published in easily accessible journals or databases. It includes things like conference proceedings that include the abstracts of the research presented at conferences, unpublished theses, and so on.

Guidelines	A WHO guideline is any document, whatever its title, that contains WHO recommendations about health interventions, whether they be clinical, public health or policy interventions. A recommendation provides information about what policy - makers, health - care providers or patients should do. It implies a choice between different interventions that have an impact on health and that have ramifications for the use of resources[1]. Systematically developed body of knowledge, integrating research evidence and descriptions of the types of other considerations needed to inform decision making about appropriate health system arrangements in specific settings (adapted from Field and Lohr 1990[200] definition of clinical guidelines). The following types of documents are not guidelines, in the sense described above, regardless how they are named: • rules, regulations, principles; • descriptions of operational procedures[201]; • lists of standards to adhere to; • presentations of options for courses of actions without any statement on their relative value. The WHO handbook for guideline development[1] explicitly excludes documents that provide information on different options for interventions without recommending any particular intervention (page 5).
Health status	The state of health of a person or population assessed with reference to morbidity, impairments, anthropological measurements, mortality, and indicators of functional status and quality of life.
Health system arrangements	The delivery, financial and governance arrangements within which clinical or public health programmes and services are provided.
Health technology assessment (HTA)	HTA is the systematic evaluation of the properties, effects and/or other impacts of health care technology. Its primary purpose is to provide objective information to support healthcare decisions and policymaking at the local, regional, national and international levels. HTA reports typically include a range of economic, social, ethical and legal considerations, as well as a review of the research evidence about the effectiveness of a technology. Some HTA reports contain a systematic review that can be applied in contexts other than the one for which the report was produced.
Heterogeneity	1. Used in a general sense to describe the variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies, or the variation in internal validity of those studies. 2. Used specifically, as statistical heterogeneity, to describe the degree of variation in the effect estimates from a set of studies. Also used to indicate the presence of variability among studies beyond the amount expected due solely to the play of chance.
Impact evaluation	An evaluation that aims to determine whether the observed changes in outcomes (or “impact”) can be attributed to a particular policy or programme.
Information	Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told; intelligence, news. Contrasted with data: that which is obtained by the processing of data (OED100).

Indicator	A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance.
Indirect evidence	Research that has not directly compared the options in which we are interested in the populations in which we are interested, or measured the important outcomes in which we are interested.
Inequity in health	A difference in health that is not only unnecessary and avoidable but, in addition, is considered unfair and unjust.
Intermediary outcome	Outcome measures that are not of direct practical importance but are believed to reflect outcomes that are important; for example, blood pressure is not directly important to patients but it is often used as an outcome in clinical trials because it is a risk factor for stroke and heart attacks. Surrogate endpoints are often physiological or biochemical markers that can be relatively quickly and easily measured, and that are taken as being predictive of important outcomes. They are often used when observation of important outcomes requires long follow-up.
Interrupted time series analysis or study	A research design that collects observations at multiple time points before and after an intervention (interruption). The design attempts to detect whether the intervention has had an effect significantly greater than the underlying trend.
Intervention	The process of intervening on people (e.g. clinical interventions), groups or entities (e.g. health policy or programme options).
Intervention group	A group of participants in a study receiving a particular policy or programme option.
Meta-analysis	The use of statistical techniques in a systematic review to integrate the results of included studies. Sometimes used as a synonym for systematic reviews, where the review includes a meta-analysis.
Meta-ethnography	A method of translating ideas, concepts and metaphors across different qualitative studies in order to synthesise their findings. The method draws on the ethnographic approach used in primary qualitative research.
Modifying factor	A factor or characteristic, such as the size of a health facility that may change or modify the effect of the proposed causal factor being studied, such as health worker motivation.
Monitoring / performance monitoring	Describes the process of systematically collecting data to inform policymakers, managers and other stakeholders whether a new policy or programme is being implemented in accordance with their expectations.
Narrative review	A summary in words (rather than numerically) of, for example, the effects of a policy or programme option. Narrative reviews are not always based on a thorough and reproducible search of the literature for studies that address the review question.
Narrative summary / synthesis	See ‘Narrative review’. The approach can be used to synthesise findings across a number of qualitative studies or to combine qualitative and quantitative evidence within a single review.

Observational study	A study in which the investigators do not seek to intervene, and simply observe the course of events. Changes or differences in one characteristic (e.g. whether or not people received the intervention of interest) are studied in relation to changes or differences in other characteristic(s) (e.g. whether or not they died), without action by the investigator. There is a greater risk of selection bias than in experimental studies. See also randomised controlled trial. (Also called non-experimental study).
Outcome	A change resulting from an intervention. In evaluations, a potential consequence of an intervention that is measured after the intervention has been implemented, that is used to assess the effects of the intervention.
PICOT (Population, Intervention, Comparison, Outcomes, Time)	The acronym, PICOT, is used to summarise the five key components of a review or research question.
Primary outcome	The outcome of greatest importance.
Primary study	'Original research' in which data are collected. The term primary study is sometimes used to distinguish it from a secondary study (re-analysis of previously collected data), meta-analysis, and other ways of combining studies (such as economic analysis and decision analysis). (Also called original study).
Process evaluation	Process evaluations explore the delivery of a process or programme and the mechanisms underlying its effects. They verify what the policy or programme is and whether or not it is delivered as intended to the target recipients.
Process indicator	An indicator for actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilised to produce specific outputs (see 'Indicator'). Processes may also be referred to as 'activities' in the results chain.
Publication bias	A bias caused by only a subset of all the relevant data being available. The publication of research can depend on the nature and direction of the study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (e.g. only outcomes or sub-groups where a statistically significant difference was found.
Qualitative comparative analysis	A method for synthesising findings from a number of qualitative studies or to combine qualitative and quantitative evidence within a single review.
Qualitative study	Qualitative approaches attempt to describe and interpret human phenomena rather than to measure these. These methods focus on finding answers to questions centred on social experience, including the values and perceptions of individuals and groups and how they experience the world around them, including health care.
Randomised controlled trial (RCT)	An experiment in which two or more interventions, possibly including a control intervention or no intervention, are compared by being randomly allocated to participants.
Realist review/synthesis	A theory-based method for synthesising findings from a number of qualitative studies.
Relative reduction	See relative effectiveness.

Retrospective analysis/study	An analysis or study planned and conducted after the dataset has already been collected. For example, routinely collected data may be analysed retrospectively to evaluate the effects of a new programme.
Risk factor	An aspect of a person's condition, lifestyle or environment that affects the probability of occurrence of a disease. For example, cigarette smoking is a risk factor for lung cancer.
Routine data	Data or information collected as part of normal health service management, monitoring and evaluation. This may include information on the prevalence of diseases, on healthcare utilisation, or on service costs.
Stakeholder	A person, group or organisation that has a legitimate interest in or can be affected by a health policy or programme.
Statistical pooling	The use of quantitative, statistical methods to combine the findings of a number of studies of the effects of programme or policy options. This is also referred to as meta-analysis and may be part of a systematic review.
Statistical significance	The likelihood that a finding or a result is caused by something other than just chance.
Subgroup analysis	An analysis in which the intervention effect is evaluated in a defined subset of the participants in a study or systematic review, or in complementary subsets, such as by sex or in age categories.
Surrogate outcome	Outcome measures that are not of direct practical importance but are believed to reflect outcomes that are important; for example, blood pressure is not directly important to patients but it is often used as an outcome in clinical trials because it is a risk factor for stroke and heart attacks. Surrogate endpoints are often physiological or biochemical markers that can be relatively quickly and easily measured, and that are taken as being predictive of important outcomes. They are often used when observation of important outcomes requires long follow-up.
Systematic review	Summaries of research evidence that address a clearly formulated question using systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review.
Trial	The term ‘trial’ is sometimes used to refer to randomised controlled trials (RCTs). The term may also be used to refer to quasi-randomised trials (e.g. where alternation (every other person or group) is used to allocate people, rather than randomisation) (Also called controlled trial).
Validity (related to indicators)	The extent to which an indicator accurately measures what it purports to measure.

Annex 3. Institutional statements: research, health systems, evidence and policy

A-1.1 The Mexico statement on health research. Knowledge for better health: strengthening health systems. Mexico city - 2004[202]

"National governments to establish sustainable programmes to support evidence-based public health and health care delivery systems, and evidence-based health related policies".

A-1.2 World Health Assembly Resolution WHA58.34 - 2005[203]

"To establish or strengthen mechanisms to transfer knowledge in support of evidence-based public health and health-care delivery systems, and evidence-based health-related policies".

"To assist in the development of more effective mechanisms to bridge the divide between ways in which knowledge is generated and ways in which it is used, including the transformation of health-research findings into policy and practice".

A-1.3 The Bamako call to action on research for health - 2008[204]

"To improve capacity in institutions, ministries, and throughout systems for the implementation of research policies, including: identifying national research priorities; […]; integrating research for health within health systems; translating research into action; and evaluating the impact of research for health".

"To promote knowledge translation and exchange through the application of effective and safe interventions, evidence-informed policies, policy-informed research, and publication and effective dissemination of research results, including to the public, taking into consideration the diversity of languages and advances in information technology".

"To invest at least 5% of development assistance funds earmarked for the health sector in research, including support to knowledge translation and evaluation as part of the research process, and to pursue innovative financing mechanisms for research for health".

A-1.4 WHO Research policy: the Guidelines Review Committee[205]

"The GRC was established by the Director General in 2007 to ensure that WHO guidelines are of a high methodological quality and are developed through a transparent, evidence-based decision-making process".

Annex 4. Flowcharts of the WHO guidelines production processes.

Figure A - 2. WHO guidelines production process.

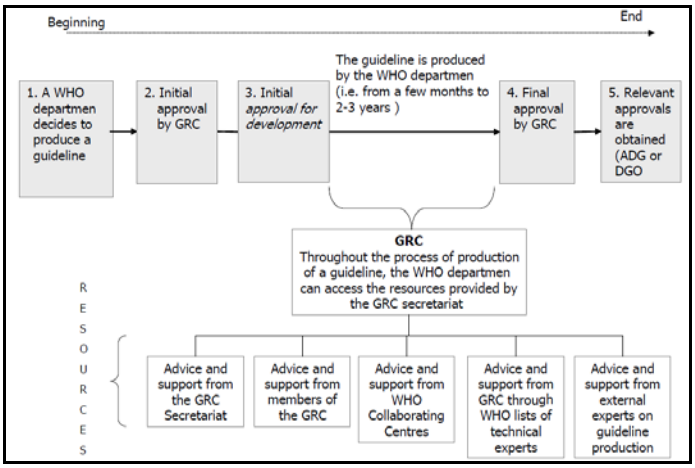


Figure A - 3. Initial approval for WHO guidelines.

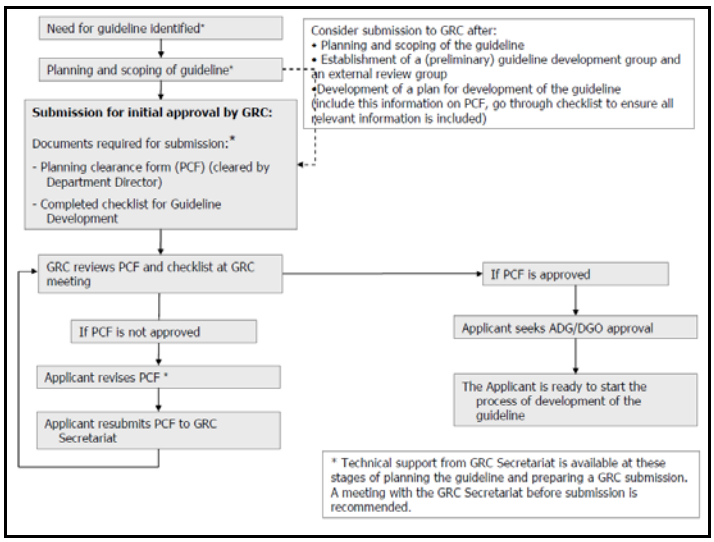
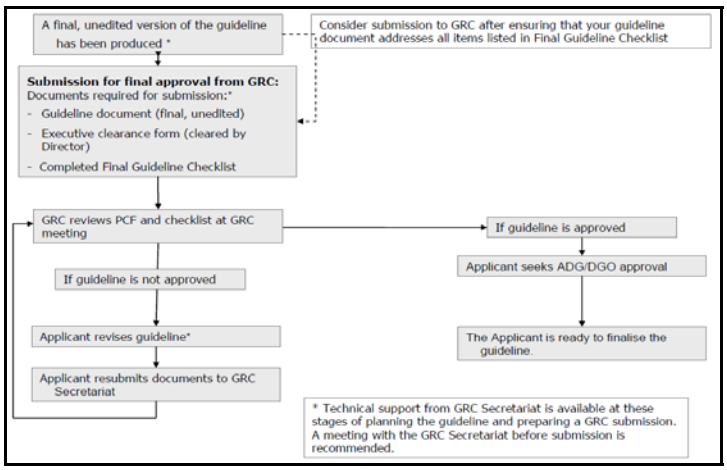


Figure A - 4. Final approval for WHO guidelines.



Annex 5. List of interventions relevant for health systems.

A-1.5 Cochrane Effective Practice and Organisation of Care group[59]

- Reviews of specific types of interventions
  - Continuing education and quality assurance
  - Distribution of educational materials
  - Educational meetings (including lectures, workshops and traineeships)
  - Local consensus processes
  - Educational outreach visits
  - Local opinion leaders
  - Patient mediated interventions
  - Audit and feedback
  - Reminders (including computerised decision support systems)
  - Marketing
  - Mass media
  - Other
  - Financial interventions
  - Provider oriented
  - Fee-for-service
  - Prepaid (no other description)
  - Capitation
  - Provider salaried service
  - Prospective payment
  - Provider incentives
  - Institution incentives
  - Provider grant/allowance
  - Institution grant/allowance
  - Provider penalty
  - Institution penalty
  - Formulary
  - Patient oriented
  - Organisational interventions
  - Provider oriented
  - Patient oriented
  - Structural interventions
  - Regulatory interventions
  - Changes in medical liability
  - Management of patient complaints
  - Peer review
  - Licensure
- Reviews of interventions to improve specific types of practice
  - Clinical preventive services
  - Diagnosis
  - Test ordering
  - Referrals
  - Procedures
  - General management of a problem
  - Patient education/advice

- Professional-patient communication
- Record keeping
- Financial (resource use)
- Discharge planning
- Patient outcomes

- Broad overviews (summaries to provide guidance for different target audiences, such as clinicians, consumers, policy makers, guideline developers, educators; e.g. strategies to introduce clinical practice guidelines)

A-1.6 Canadian Agency for Drugs and Technologies in Health[60]

- Professional: interventions that target professionals directly, aiming to improve practice.
  - Distribution of Educational Materials
  - Educational Meetings
  - Local Consensus Process
  - Educational Outreach Visits
  - Local Opinion Leaders
  - Patient-Mediated
  - Audit and Feedback
  - Reminders - General
  - Reminders - Computer Decision Support Systems (drug dosing)
  - Reminders - Computer Physician Order Entry
  - Tailored Interventions
  - Mass Media
  - Professional - other
  - Multifaceted New Intervention
- Consumer: interventions that affect drug use by and prescribing for consumers.
  - Providing information or education
  - Supporting behaviour change
  - Acquiring skills and competencies
  - Support
  - Facilitating communication and decision making
  - Minimising risks or harms
  - Improving quality
  - Consumer system participation
- Organizational: interventions that involve a change in the structure or delivery of health care.
  - Revision of Professional Roles - Pharmacy
  - Revision of Professional Roles - Nursing
  - Revision of Professional Roles - General New Intervention
  - Quality Improvement
  - Changes in medical records systems
  - Changes in physical structure, facilities and equipment
  - Changes in scope and nature of benefits and services
  - Changes to settings/site of delivery service
  - Clinical multidisciplinary teams
  - Continuity of care
  - Skill mix changes
  - Communication and case discussion between distant health professionals

- Organisational (Patient oriented interventions)
- Formal integration of services
- Satisfaction of providers with the conditions of work and the material and psychological rewards
- Ownership, accreditation, and affiliation status of hospitals and other facilities
- Organisational (Provider) - Other
- Organisational (Structural) - Other

- Financial: interventions that change professional reimbursement, incentives and penalties.
  - Financial
  - Formulary
- Regulatory: interventions that aim to change health services delivery by regulation or law.
  - Regulatory

A-1.7 Taxonomy of governance, financial and delivery arrangements for primary health[61]

Adapted from Lavis et al[206].

- Governance arrangements
  - Policy authority; e.g. who makes policy decisions about what primary health care encompasses (such as whether such decisions are centralised or decentralised)
  - Organisational authority; e.g. who owns and manages primary health-care clinics (such as whether private for-profit clinics exist)
  - Commercial authority; e.g. who can sell and dispense antibiotics in primary health care and how they are regulated
  - Professional authority; e.g. who is licensed to deliver primary health-care services; how is their scope of practice determined; and how they are accredited
  - Consumer and stakeholder involvement—who from outside government is invited to participate in primary health-care policy-making processes and how are their views taken into consideration
- Financial arrangements
  - Financing; e.g. how revenue is raised for core primary health-care programmes and services (such as through community-based insurance schemes)
  - Funding; e.g. how primary health-care clinics are paid for the programmes and services they provide (such as through global budgets)
  - Remuneration; e.g. how primary health-care providers are remunerated (such as via capitation)
  - Financial incentives; e.g. whether primary health-care patients are paid to adhere to care plans
  - Resource allocation; e.g. whether drug formularies are used to decide which medications primary health-care patients receive for free
- Delivery arrangements
  - To whom care is provided and the efforts that are made to reach them (such as interventions to ensure culturally appropriate primary health care)
  - By whom care is provided (such as primary health-care providers working autonomously vs as part of multidisciplinary teams)

- Where care is provided; e.g. whether primary health care is delivered in the home or community health facilities
- With what information and communication technology is care provided; e.g. whether primary health care record systems are conducive to providing continuity of care
- How the quality and safety of care is monitored; e.g. whether primary health-care focused quality-monitoring systems are in place

Annex 6. The COGS checklist for reporting clinical practice guidelines.

Table A - 10-1. The COGS checklist for reporting clinical practice guidelines.

Item	Description
1. Overview material	Provide a structured abstract that includes the guideline's release date, status (original, revised, updated), and print and electronic sources.
2. Focus	Describe the primary disease / condition and intervention / service / technology that the guideline addresses. Indicate any alternative preventive, diagnostic or therapeutic interventions that were considered during development.
3. Goal	Describe the goal that following the guideline is expected to achieve, including the rationale for development of a guideline on this topic.
4. Users/setting	Describe the intended users of the guideline (e.g., provider types, patients) and the settings in which the guideline intended to be used.
5. Target population	Describe the patient population eligible for guideline recommendations and list any exclusion criteria.
6. Developer	Identify the organization(s) responsible for guideline development and the names / credentials / potential conflicts of interest of individuals involved in the guideline's development.
7. Funding source/sponsor	Identify the funding source/sponsor and describe its role in developing and/or reporting the guideline. Disclose potential conflict of interest.
8. Evidence collection	Describe the methods used to search the scientific literature, including the range of dates and databases searched, and criteria applied to filter the retrieved evidence.
9. Recommendation grading criteria	Describe the criteria used to rate the quality of evidence that supports the recommendations and the system for describing the strength of the recommendations. Recommendation strength communicates the importance of adherence to a recommendation and is based on both the quality of the evidence and the magnitude of anticipated benefits or harms.
10. Method for synthesizing evidence	Describe how evidence was used to create recommendations, e.g. evidence tables, meta-analysis, decision analysis.
11. Prerelease review	Describe how the guideline developer reviewed and/or tested the guidelines prior to release.
12. Update plan	State whether or not there is a plan to update the guideline and, if applicable, and expiration date for this version of the guideline.
13. Definitions	Define unfamiliar terms and those critical to correct application of the guideline that might be subject to misinterpretation.
14. Recommendations and rationale	State the recommended action precisely and the specific circumstances under which to perform it. Justify each recommendation by describing the linkage between the recommendation and its supporting evidence. Indicate the quality of evidence and the recommendation strength, based on the criteria described in 9.
15. Potential benefits and harms	Describe anticipated benefits and potential risks associated with implementation of guideline recommendations.
16. Patient preferences	Describe the role of patient preferences when a recommendation involves a substantial element of personal choice or values.
17. Algorithm	Provide (when appropriate) a graphical description of the stages and decisions in clinical care described by the guideline.
18. Implementation considerations	Describe anticipated barriers to application of the recommendations. Provide reference to any auxiliary documents for providers or patients that are intended to facilitate implementation. Suggest review criteria for measuring changes in care when the guideline is implemented.

Annex 7. Table to manage the list of people and teams involved in guidance development

Table A - 10-2. Members of guidance development groups.

PEOPLE AND TEAMS FOR GUIDANCE DEVELOPMENT														
Group	**Role**	**Panel**	**Gender**	**Title**	**First name**	**Family name**	**Affiliation**	**Organisation**	**Country**	**e-mail**	**phone**	**TOR**	**DOI**	**CV**
Guidance steering group	**Coordinator**	**yes**										Pending	Pending	Pending
	Team member											Pending	Cleared	Pending
	Team member											Signed	Cleared	Pending
												Pending	Not needed	Pending
Guidance development group	**Chair**											Pending	Pending	Pending
	Member											Pending	Pending	Pending
	Member											Pending	Pending	Pending
												Pending	Pending	Pending
Panel	**Facilitator**	**yes**										Pending	Pending	Pending
	Additional member	**yes**										Pending	Pending	Pending
	Additional member	**yes**										Pending	Pending	Pending
												Pending	Cleared	Pending
Technical support	**Literature search**											Pending	Pending	Not needed
	Researcher											Pending	Pending	Available
	Researcher											Pending	Pending	Pending
	Policy											Pending	Pending	Pending
	Communication											Pending	Pending	Available
	Design											Pending	Pending	Not needed
	Translator 1											Pending	Pending	Pending
	Translator 2											Pending	Pending	Pending
												Pending	Pending	Pending
External reviewers	**In favour**											Pending	Pending	Pending
	In favour											Pending	Pending	Pending
	In favour											Pending	Pending	Pending
	Neutral											Pending	Pending	Pending
	Neutral											Pending	Pending	Pending
	Neutral											Pending	Pending	Pending
	Critical											Pending	Pending	Pending
	Critical											Pending	Pending	Pending
	Critical											Pending	Pending	Pending
												Pending	Pending	Pending

CV: curriculum vitae; DOI: declaration of interests; TOR: terms of reference.

Annex 8. Searching for health systems guidance

Table A - 10-3. Sources, search strategy and hits to retrieve health systems guidance.

Sources	Search	Hits
Pubmed	(guideline*[ti] OR guidance*[ti] OR recommend*[ti]) AND ((health service*) OR (health system*)) AND polic*	508
VHL Cochrane	(Ti:guideline OR Ti:guidelines OR Ti:guidance OR Ti:recommendation OR Ti:recommendations) AND (policy OR policies)	301
VHL Health Science Descriptors		5
VHL HIL		30
VHL IBECS		15
VHL LILACS		193
VHL PAHO		104
VHL Specialised		122
VHL WHOLIS		117
Total		**1,395**
Duplicates		26
To assess		**1,369**

We searched (November 2011) in the Virtual Health Library[1] and Medline[2].

Virtual Health Library DeCS/MeSH terms for 'Guideline': /guidelines, Guideline, Guideline Adherence, Guidelines as Topic, Guidelines for Health Planning, Guidelines for Radiation Safety, Guidelines for Radiological Safety, Health Planning Guidelines, Manuals and Guidelines for Research Management, Practice Guideline, Practice Guidelines as Topic, Radiation Safety Guidelines, Radiological Safety Guidelines.

[1] Virtual Health Llibrary (VHL). http://regional.bvsalud.org/php/decsws.php (accessed 11/2010).
[2] NCBI. http://www.ncbi.nlm.nih.gov/mesh?term=guideline (accessed 11/2010).

Syntax: "Practice Guidelines as Topic" OR "Guidelines as Topic" OR "Health Planning Guidelines". Hits:

- General Health Sciences (77264 Result)
    - LILACS - Latin American and Caribbean Health Sciences (782)
    - IBECS (359)
    - MEDLINE - International Database for Medical Literature (76123)
- Specialized Areas (170 Result)
    - CidSaúde - Literatura sobre Cidades/Municípios Saudáveis (7)
    - DESASTRES - Disaster Documentation Center Collection (11)
    - HOMEOINDEX - Homeopathy Brazilian Bibliography (1)
    - LEYES - Latin American and Caribbean Basic Health Legislation (0)
    - MEDCARIB - Caribbean Health Sciences Literature (0)
    - REPIDISCA - Sanitary Engineering and Environmental Sciences (151)
- International Agencies (1766 Result)
    - PAHO - PAHO HQ Library Catalog (85)
    - WHOLIS - World Health Organization Library Information System (1681)
- HIL- Health Information Locator (0 Result)
- Document Access (0 Result)
- Events directory (0 Result)
- VHL network directory (0 Result)

Pubmed, search MeSH for ‘guideline’: Guideline [Publication Type], Guidelines as Topic, Guideline Adherence, Practice Guideline [Publication Type], Health Planning Guidelines, Double Effect Principle, Advance Directive Adherence.

Syntax: “Guideline "[Publication Type] AND (("Guidelines as Topic"[Mesh] OR "Health Planning Guidelines"[Mesh])). Hits: 967.

References from WHOLIS (VHL) and Medline were downloaded.

Annex 9. Sources of evidence

Adapted from NICE[71] with inputs from other documents[1].

Core databases

- AMED (Allied and Complementary Medicine)
- ASSIA (Applied Social Science Index and Abstracts)
- British Nursing Index
- CINAHL (Cumulative Index of Nursing and Allied Health Literature)
- Cochrane Central Register of Controlled Trials
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness (DARE; 'other reviews' in Cochrane Library)
- EMBASE
- Global Health Library (http://www.who.int/ghl/en/)
- HMIC (or Kings Fund catalogue and DH data)
- MEDLINE
- UK Clinical Research Network Portfolio Database
- PsycINFO
- Sociological Abstracts
- Social Policy and Practice
- Social Science Citation Index

Additional topic specific databases for consideration

- ABI Inform (business/workplace issues)
- Ageline (older people)
- Campbell Collaboration reviews
- Enviroline (environment)
- EPPI Centre databases (qualitative reviews)
- ERIC (education) Methods for the development of NICE public health guidance (second edition) 189
- McMaster University. Health Systems Evidence. http://www.healthsystemsevidence.org/.
- McMaster University. The Program in Policy Decision-Making/Canadian Cochrane Network and Centre (PPD/CCNC). http://library.mcmaster.ca/articles/ppdccnc-database.
- PAIS International (Public Affairs Information Service)
- SportDiscus (sport)
- Social Care Online (SCIE) (social care)
- Transportation Research Information Services (TRIS) (transport)
- US National Library of Medicine. National Institutes of Health. National Information Center on Health Services Research and Health Care Technology (NICHSR). http://www.nlm.nih.gov/nichsr/.

Core websites

- NICE website. In addition, former Health Development Agency documents should be searched at: www.nice.org.uk/page.aspx?o=hda.publications
- NHS. National Institute for Health Research. http://www.crd.york.ac.uk/cms2web/SearchPage.asp.
- Public health observatories.
- Relevant National Library for Health Specialist Library.
- Joseph Rowntree Foundation.

Annex 10. Systematic review critical appraisal tools

From the Institute of Medicine[95]:

Standards for Reporting Systematic Reviews

STANDARD 5.1
Prepare final report using a structured format

5.1.1 Include a report title

5.1.2 Include an abstract

5.1.3 Include an executive summary

5.1.4 Include a summary written for the lay public

5.1.5 Include an introduction (rationale and objectives)

5.1.6 Include a methods section. Describe the following:

- Research protocol
- Eligibility criteria (criteria for including and excluding studies in the systematic review)
- Analytic framework and key questions
- Databases and other information sources used to identify relevant studies
- Search strategy
- Study selection process
- Data extraction process
- Methods for handling missing information
- Information to be extracted from included studies
- Methods to appraise the quality of individual studies
- Summary measures of effect size (e.g., risk ratio, difference in means)
- Rationale for pooling (or not pooling) results of included studies
- Methods of synthesizing the evidence (qualitative and meta-analysis)
- Additional analyses, if done, indicating which were prespecified

5.1.7 Include a results section. Organize the presentation of results around key questions. Describe the following (repeat for each key question):

- Study selection process
- List of excluded studies and reasons for their exclusion
- Appraisal of individual studies' quality
- Qualitative synthesis
- Meta-analysis of results, if performed (explain rationale for doing one)
- Additional analyses, if done, indicating which were prespecified
- Tables and figures

5.1.8 Include a discussion section. Include the following:

- Summary of the evidence
- Strengths and limitations of the systematic review
- Conclusions for each key questions
- Gaps in evidence
- Future research needs

5.1.9 Include a section describing funding sources and COI

STANDARD 5.2
Peer review the draft report

5.2.1 Use a third party to manage the peer review process

5.2.2 Provide a public comment period for the report and publicly report on disposition of comments

STANDARD 5.3
Publish the final report in a manner that ensures free public access

Adapted from the Centre for Evidence Based Medicine[207].

SYSTEMATIC REVIEW: Are the results of the review valid?

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

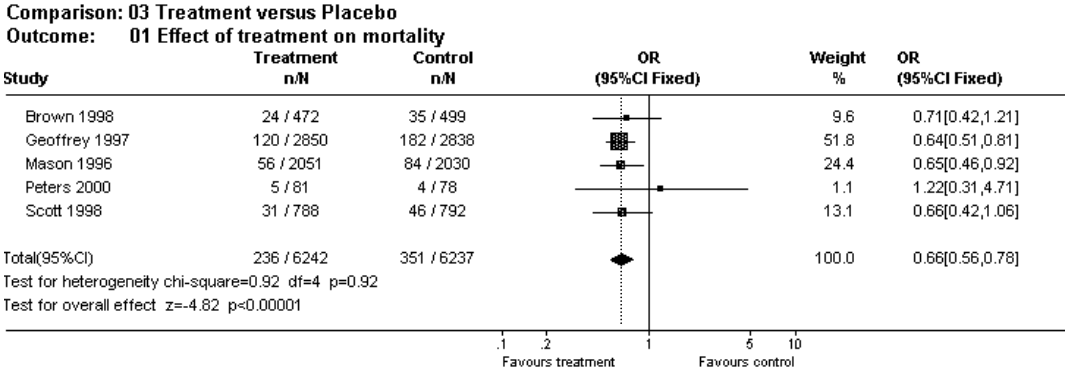
This paper: Yes □ No □ Unclear □

Comment:

What were the results?

How are the results presented?

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



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

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

Exploring heterogeneity

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

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

Annex 11. Detail of GRADE criteria

Adapted form the description appeared in the BMJ 2004[208].

Relative importance of outcomes

Only important outcomes should be included in evidence profiles. The included outcomes should be classified as critical or important (but not critical) to a decision). This greatly depends on the perspective: "it is essential to take cultural diversity into account when deciding on relative importance of outcomes, particularly when developing recommendations for an international audience".

1. Type of evidence (design):

Randomised trial = high
Observational study = low
Any other evidence = very low

2. Decrease grade if:

- **Quality**: serious (-1) or very serious (-2) limitation to study quality
- **Consistency**: important inconsistency (-1)
- **Directness**: some (-1) or major (-2) uncertainty about directness in comparison, population, intervention, comparator and outcome.
- Imprecise or sparse data (-1)
- High probability of reporting bias (-1)

3. Increase grade if:

- Strong evidence of association—significant relative risk of > 2 (< 0.5) based on consistent evidence from two or more observational studies, with no plausible confounders (+1)
- Very strong evidence of association—significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity (+2)
- Evidence of a dose response gradient (+1)
- All plausible confounders would have reduced the effect (+1)

4. Overall quality of evidence (across outcomes)

The overall quality of evidence should be judged across outcomes based on the lowest quality of evidence for any of the critical outcomes.

- 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.

Strength of recommendations

- Balance between desirable and undesirable effects
    - Importance of outcomes
    - Baseline risk of outcomes
    - Relative and absolute effect of an intervention
    - Precision of the estimates of the effects
    - Cost
- Quality of evidence
- Values and preferences

- Costs

5. Balance of benefits and harms

The balance of benefits and harms should be classified as net benefits, trade-offs, uncertain trade-offs, or no net benefits based on the important health benefits and harms

- Net benefits = the intervention clearly does more good than harm.
- Trade-offs = there are important trade-offs between the benefits and harms.
- Uncertain trade-offs = it is not clear whether the intervention does more good than harm.
- No net benefits = the intervention clearly does not do more good than harm.

6. Balance of net benefits and costs

Are incremental health benefits worth the costs? Because resources are always limited, it is important to consider costs (resource utilisation) when making a recommendation

7. Strength of recommendation

Recommendations should be formulated to reflect their strength—that is, the extent to which one can be confident that adherence will do more good than harm

- The trade-offs take into account the estimated size of the effect for the main outcomes, the confidence limits around those estimates, and the relative value placed on each outcome
- The quality of the evidence
- Translation of the evidence into practice in a specific setting, taking into consideration important factors that could be expected to modify the size of the expected effects, such as proximity to a hospital or availability of necessary expertise
- Uncertainty about baseline risk for the population of interest.

Categories for recommendations:

- “Do it” or “don't do it”: indicating a judgment that most well informed people would make;
- “Probably do it” or “probably don't do it”: indicating a judgment that a majority of well informed people would make but a substantial minority would not.

Annex 12. Evidence dissemination

Table A - 10-4. Summary of key features of evidence / knowledge translation / dissemination initiatives.

Initiative	Thematic focus	Evidence	Geographic focus	Research synthesis and Knowledge translation
AHPSR[209]	Health policy and systems	Several types	All LMICs	Briefing notes; research issues; links to SUPPORT summaries
Bandolier[210]	Potentially vast: family health, infectious diseases, management, vaccines…	Systematic reviews	UK	'bullet points of those things that work and those that do not'
Canadian HSRF[211]	Human resources, primary health care organisation, management and quality of care; decision-making and public engagement	Importance of high quality research is emphasised	Canada	Knowledge summaries with different focus[212,213,214 215]
Cochrane Library[216]	EPOC review group focuses on health systems questions	Systematic reviews	global	Systematic reviews include plain language summaries
EVIPNet[217]	Health systems, including what services and programmes to offer or cover, how to deliver those services; financial arrangements; governance arrangements; and how to bring about change	Systematic reviews	LMICs	EvIPNet policy briefs (research syntheses in user-friendly format)
HRCS[218]	'Health research'	Not stated	Kenya and Malawi	Available info does not mention summaries
McMaster University[219]	Health systems	Systematic reviews	Canadian priorities (not limited to)	Policy briefs, systematic reviews, overviews
REACH-PI[220]	Health policy and systems	Not stated	Kenya, Tanzania, Uganda	Intended to do knowledge brokering
SUPPORT[221]	Maternal and child health; health systems questions relevant to MCH	Systematic reviews	All LMICs	Clearly stated method and transparent process
WHO RHL[222]	Reproductive health, mainly clinical topics	Cochrane reviews	Global, esp LMICs	Effectiveness summaries, mainly on clinical topics
ZamFoHR[223]	Health research	Not stated	Zambia	Sites says Zamfohr does research syntheses- but none are available at present

Adapted from Kelly 2009[99]. LMIC: Low- and middle- income countries.

Annex 13. List of research topics

References

[1] WHO. WHO handbook for guideline development. March 2010.

[2] Bosch-Capblanch X. Analytical framework for health systems guidance. Health Systems handbook project. Swiss TPH. September 2010.

[3] How to find clinical practice guidelines. http://www.youtube.com/watch?v=63rHFBpzEsY.

[4] Canadian partnership against cancer. Critical Appraisal of Evidence. http://www.youtube.com/watch?v=KL_yC0RYc6Y&feature=related.

[5] Biesma RG, Brugha R, Harmer A, Walsh A, Spicer N, Walt G. The effects of global health initiatives on country health systems: a review of the evidence from HIV/AIDS control. Health Policy and Planning 2009. 24: 239–52.

[6] Travis P, Bennett S, Haines A, Pang T, Bhutta Z, Hyder AA, Pielemeier AR, Mills A, Evans T. Overcoming health-systems constraints to achieve the Millennium Development Goals. Lancet 2004; 364: 900-6.

[7] Victora CG, Fenn B, Bryce J, Kirkwood BR. Co-coverage of preventive interventions and implications for child-survival strategies: evidence from national surveys. Lancet 2005; 366: 1460-6.

[8] Sundewall J, Swanson RC, Betigeri A, Sanders D, Collins TE, Shakarishvili G, Brugha R. Health-systems strengthening: current and future activities. Lancet 2010; DOI:10.1016/S0140-6736(10)60679-4.

[9] Strengthening Health Systems. http://ghsia.wordpress.com/.

[10] WHO. Health systems strengthening. http://www.who.int/healthsystems/en/ (accessed August 2010).

[11] Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. BMC Health Research Policy and Systems 2006, 4:13 doi:10.1186/1478-4505-4-13.

[12] Raine R, Sanderson C, Black N. Developing clinical practice guidelines: a challenge to current methods. British Medical Journal 2005; 331: 631-3.

[13] Chowdhury M, Rosenfeld A (coordinators). Interim Report of Task Force 4 on Child Health and Maternal Health. MillenniumProject 2004.

[14] Hardee K, Feranil I, Boezwinkle J, Clark B. The policy cycle: a Framework for Analyzing the Components of Family Planning, Reproductive Health, Maternal Health, and HIV/AIDS Policies. Policy working paper series 11. June 2004.

[15] WHO. Primary Health Care. Now more than ever. The World Health Report 2008.

[16] Cooper HM. Scientific guidelines for conducting integrative research reviews. Review of Educational Research 1982; 52: 291-302.

[17] Lavis JN, Wilson MG, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 4: Using research evidence to clarify a problem. Health Research Policy and Systems 2009, 7(Suppl 1):S4 doi:10.1186/1478-4505-7-S1-S4.

[18] Quality Assurance project. Health care and workforce improvement. http://www.qaproject.org/methods/resources.html#QAtools.

[19] WHO. Health workforce. Strengthening health workforce to strengthen health systems. http://www.who.int/hrh/resources/strengthening_hw/en/index.html.

[20] WHO. Increasing access to health workers in remote and rural areas through improved retention. Global policy recommendations. WHO 2010. http://whqlibdoc.who.int/publications/2010/9789241564014_eng.pdf.

[21] WHO. Everybody's business: Strengthening Health Systems to Improve Health Outcomes: WHO's framework for Action. Geneva 2007.

[22] De Savigny D, Adam T (editors). Systems thinking for Health Systems Strengthening. AHPSR 2009.

[23] HMN. Framework and standards for country Health Information Systems. Second Edition. WHO. 2008.

[24] Lewin S, Oxman AD, Lavis JN, Fretheim A, Garcia Marti S, Munabi-Babigumira S. SUPPORT Tools for evidence-informed Policymaking in health 11: Finding and using evidence about local conditions. Health Research Policy and Systems 2009, 7(Suppl 1):S11 doi:10.1186/1478-4505-7-S1-S11.

[25] IMF. Data Quality Assessment framework – Generic framework (July 2003 framework). http://dsbb.imf.org/pages/dqrs/DQAF.aspx

[26] WHO-HSS. Increasing access to health workers in remote and rural areas through improved retention. Background paper for the first expert meeting to develop evidence-based recommendations to increase access to health workers in remote and rural areas through improved retention Geneva, 2-4 February, 2009. Draft. Health Systems and Services cluster - February 2009. http://who.int/hrh/migration/background_paper.pdf.

[27] STROBE Statement. checklist of items that should be included in reports of observational studies. 2007. http://www.strobe-statement.org/index.php?id=available-checklists.

[28] Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. Health Research Policy and Systems. 2006, 4:14 doi:10.1186/1478-4505-4-14.

[29] Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. Health Research Policy and Systems 2006, 4:14 doi:10.1186/1478-4505-4-14.

[30] US Department of Health and Human Services. Agency for Health Care Research and Quality. National Guideline Clearinghouse. http://www.guideline.gov/browse/by-topic.aspx.

[31] McMaster University. Health Systems evidence. http://www.healthsystemsevidence.org/.

[32] NHS – National Institute for Health and Clinical Excellence. Order NICE guidance. http://guidance.nice.org.uk/.

[33] EVIPNet. Evidence-informed policy network. http://www.evipnet.org/modules/dia/index.php?where=POLICYBRIEF#.

[34] WHO. Research policy. http://www.who.int/rpc/en/.

[35] HEN. Health Evidence Network. Publications. http://www.euro.who.int/en/what-we-do/data-and-evidence/health-evidence-network-hen/publications.

[36] Virtual Health Library. http://regional.bvsalud.org/php/index.php.

[37] Trip database. www.tripdatabase.com.

[38] Australian Government. National Health and Medical Research Council. NHMRC Guidelines and Publications. http://www.nhmrc.gov.au/guidelines.

[39] US National Library of Medicine. National Information Center on Health Services Research and Health Care Technology (NICHSR). http://www.nlm.nih.gov/nichsr/hedges/search.html.

[40] University of Massachusetts Medical School. Evidence based guidelines. http://library.umassmed.edu/ebpph/guidelines.cfm

[41] WHO. Library and Information Networks for Knowledge Databases. WHOLIS. http://dosei.who.int/uhtbin/cgisirsi/gsS9MNeCnC/65030034/38/1/X/BLASTOFF.

[42] EPOC. Guide for review authors. 3 – Developing a protocol.

[43] EPOC. Cochrane Effective Practice and Organisation of Care Group. EPOC resources for review authors. http://epoc.cochrane.org/epoc-resources-review-authors.

[44] EBSCO Publishing. CINAHL Databases. http://www.ebscohost.com/cinahl/.

[45] Dialog DataStar. DHSS - DH-Data; Health Admin., Medical Toxicology & Env. Health. http://ds.datastarweb.com/ds/products/datastar/sheets/dhss.htm.

[46] EMBASE biomedical answers. http://www.embase.com/home.

[47] rdrb. Research and Development Resource Base. http://www.rdrb.utoronto.ca/.

[48] OpenSigle. http://opensigle.inist.fr/.

[49] Higgins JPT, Green S eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.2. The Cochrane Collaboration [updated September 2009]. http://www.cochrane-handbook.org/.

[50] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, Littlejohns P, Makarski J, Zitzelsberger L, for the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in health care. CMAJ 2010. DOI:10.1503/cmaj.090449.

[51] NHMRC. NHMRC standards and procedures for externally developed guidelines. National Health and Medical Research Council. Australia 2007.

[52] ADAPTE. http://www.adapte.org/www/index.php.

[53] Lomas J, Culyer T, McCutcheon C, McAuley L, Law S. Conceptualising and combining evidence for health systems guidance. Final report. CHSRF 2005.

[54] Schünemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R, et al. Transparent Development of the WHO Rapid Advice Guidelines. PLoS Med 2007; 4(5): e119. doi:10.1371/journal.pmed.0040119.

[55] UK Civil Service. Rapid Evidence Assessment Toolkit. http://www.civilservice.gov.uk/my-civil-service/networks/professional/gsr/resources/gsr-rapid-evidence-assessment-toolkit.aspx

[56] WHO-Commission on Social Determinants on Health. Closing the gap in a generation Health equity through action on the social determinants of health. WHO 2008. http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf.

[57] Evans T, Brown H. Road traffic crashes: operationalising equity in the context of health sector reform. Injury Control and Safety Promotion 2003; 10: 11-2.

[58] Lagarde M, Palmer N. The impact of user fees on access to health services in low- and middle-income countries. Cochrane Database of Systematic Reviews 2011, Issue 4. Art. No.: CD009094. DOI: 10.1002/14651858.CD009094.

[59] Cochrane Effective Practice and Organisation of Care Group. EPOC Reviews. http://epoc.cochrane.org/epoc-reviews.

[60] Canadian Agency for Drugs and Technologies in Health. Rx for change. http://www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions/browse.

[61] Lewin S, Lavis JN, Oxman AD, Bastías G, Chopra M, Ciapponi A, Flottorp S, García Martí S, Pantoja T, Rada G, Souza N, Treweek S, Wiysonge CS, Haines A. Alma-Ata: Rebirth and Revision 2. Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews. Lancet 2008; 372: 928–39

[62] Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database of Systematic Reviews 2008, Issue 2. Art. No.: CD000011. DOI: 10.1002/14651858.CD000011.pub3.

[63] Donabedian A. Evaluating the quality of medical care. The Milibank Memorial Fund Quarterly 1966; 4(3): 166-203-.

[64] UNDP. Handbook on planning, monitoring and evaluating for development results. United Nations Development Programme. 2009. http://stone.undp.org/undpweb/eo/evalnet/Handbook2/documents/english/pme-handbook.pdf.

[65] Tanahashi T. Health service coverage and its evaluation. Bulletin of the World Health Organisation 1978. 56(2): 295-303.

[66] Light RJ, Pillemer DB. Summing Up: The Science of Reviewing Research. Cambridge (MA): Harvard University Press, 1984. Quoted in Cochrane Handbook for Systematic Reviews of Interventions.

[67] Eccles MP, Mittman BS. Welcome to Implementation Science. Implementation Science 2006, 1:1 doi:10.1186/1748-5908-1-1.

[68] NHMRC. Using socioeconomic evidence in clinical practice guidelines. National Health and Medical Research Council. Commonwealth of Australia 2003. http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/cp89.pdf.

[69] Deurenberg R, Vlayen J, Guillo S, Oliver TK, Fervers B, Burgers J, on behalf of the SEARCH group. Standardization of search methods for guideline development: an international survey of evidence - based guideline development groups. Health Information and Libraries Journal 2008; 25:2̃30.

[70] EVIPNet. Research synthesis. http://www.evipnet.org/php/level.php?lang=en&component=101&item=2.

[71] NHS. Methods for the development of NICE public health guidance (second edition). NHS – National Institute for Health and Clinical Excellence. April 2009.

[72] Oxman AD, Lavis JN, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 10: Taking equity into consideration when assessing the findings of a systematic review. Health Research Policy and Systems 2009, 7(Suppl 1):S10 doi:10.1186/1478-4505-7-S1-S10.

[73] EPOC. Guide for review authors. 3 – developing a protocol. Effective Practice and Organisation of Care group. http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Reviewer%20Resource%20Package%203%20-%20Developing%20a%20Protocol%20-%2003-17-2011.doc.

[74] Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. Health Promot. Int. (March 2006) 21 (1): 76-83. doi: 10.1093/heapro/dai025.

[75] Fretheim A, Munabi-Babigumira S, Oxman AD, Lavis JN, Lewin L. SUPPORT Tools for Evidence-informed Policymaking in health 6: Using research evidence to address how an option will be implemented. BMC Health Research Policy and Systems 2009, 7(Suppl 1):S6 doi:10.1186/1478-4505-7-S1-S6.

[76] Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 9: Assessing the applicability of the findings of a systematic review. Health Research Policy and Systems 2009, 7(Suppl 1):S9 doi:10.1186/1478-4505-7-S1-S9 4p.

[77] Bravata DM, McDonald KM, Shojania KG, Sundaram V, Owens DK. Challenges in Systematic Reviews: Synthesis of Topics Related to the Delivery, Organization, and Financing of Health Care. Annals of Internal Medicine 2005; 142: 1056-65.

[78] Gericke CA, Kurowski C, Ranson MK, Mills A. Intervention complexity — a conceptual framework to inform priority-setting in health. Bulletin of the World Health Organization 2005; 83: 285-93.

[79] Allotey P, Reidpath DD, Ghalib H, Pagnoni F, Skelly WC. Efficacious, effective, and embedded interventions: Implementation research in infectious disease control. BMC Public Health 2008, 8:343 doi:10.1186/1471-2458-8-343.

[80] Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. Journal of Health Services Research and Policy 2005; 10(Suppl 1): 6-20.

[81] CRD. Systematic reviews. CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination. 2009. http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm.

[82] Tugwell P, de Savigny D, Hawker G, Robinson V. Applying clinical epidemiological methods to health equity: the equity effectiveness loop. BMJ 2006; 332:358–61.

[83] Tugwell P, Robinson V, Grimshaw J, Santesso N. Systematic reviews and knowledge translation. Bulletin of the World Health Organisation 2006; 84: 643-51.

[84] NICE. The guidelines manual. NHS – National Institute for Health and Clinical Excellence. 2009.

[85] HTA. Guidelines for the Economic Evaluation of Health Technologies: Canada. Canadian Agency for Drugs and Technologies in Health 2006.

[86] Edejer TTT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. Health Research Policy and Systems 2006, 4:23 doi:10.1186/1478-4505-4-23.

[87] Centre for Reviews and Dissemination. Information resources in health economics. University of York. http://www.york.ac.uk/inst/crd/econ4.htm.

[88] Thomson Reuters. Reference Manager. http://www.refman.com/.

[89] Thomson Reuters. Endnote. http://www.endnote.com/.

[90] Thomson Reuters. Procite. http://www.procite.com/.

[91] EndNote Web. http://www.myendnoteweb.com/EndNoteWeb.html.

[92] F. Dell'Orso. Bibliography Management Software. http://www.burioni.it/forum/dellorso/bms-dasp/text/index.html.

[93] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097.

[94] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology 2007, 7:10 doi:10.1186/1471-2288-7-10.

[95] Institute of Medicine. Finding what works in health care. Standards for Systematic Reviews. 2011.

[96] Equator network. Library for health research reporting. http://www.equator-network.org/resource-centre/library-of-health-research-reporting/.

[97] Tugwell P, Petticrew M, Kristjansson E, Welch V, Ueffing E, Waters E, Brockhoff J, Bonnefoy J, Morgan A, Doohan E, Kelly MP. Assessing equity in systematic reviews: realising the recommendations of the Commission on Social Determinants of Health. BMJ 2010; 341:c4739 doi: 10.1136/bmj.c4739

[98] Lavis JN, Oxman AD, Grimshaw J, Johansen M, Boyko JA, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 7: Finding systematic reviews. BMC Health Research Policy and Systems 2009, 7(Suppl 1):S7 doi:10.1186/1478-4505-7-S1-S7.

[99] Kelly M, Bosch-Capblanch X. GAVI Health Systems Strengthening Knowledge Bank. Identification of relevant topics and mapping of existing initiatives for knowledge synthesis and translation. Report. February 2009.

[100] Oxford English Dictionary. http://www.oed.com/.

[101] Lavis JN, Wilson MG, Oxman AD, Grimshaw J, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 5: Using research evidence to frame options to address a problem. BMC Health Research Policy and Systems 2009. 7(Suppl 1):S5 doi:10.1186/1478-4505-7-S1-S5.

[102] Oliver S, Kavanagh J, Caird J, Lorenc T, Oliver K, Harden A, Thomas J, Greaves A, Oakley A (2008) Health promotion, inequalities and young people's health: a systematic review of research. London: EPPI-Centre, Social Science. Research Unit, Institute of Education, University of London.

[103] GRADE working group. http://www.gradeworkinggroup.org/toolbox/index.htm.

[104] GRADEpro. [Computer program]. Version 3.2 for Windows. Jan Brozek, Andrew Oxman, Holger Schünemann, 2008.

[105] Cochrane IMS. Download GRADEpro. http://ims.cochrane.org/revman/other-resources/gradepro/download.

[106] Balshem H, Helfanda M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt JH. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology 2011; 64 401-406.

[107] Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. BMC Health Research Policy and Systems 2006, 4:21 doi:10.1186/1478-4505-4-21.

[108] Lewin S, Munabi-Babigumira S, Glenton C, Daniels K, Bosch-Capblanch X, van Wyk BE, Odgaard-Jensen J, Johansen M, Aja GN, Zwarenstein M, Scheel IB. Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases. Cochrane Database of Systematic Reviews 2010, Issue 3. Art. No.: CD004015. DOI: 10.1002/14651858.CD004015.pub3.

[109] Hanson K, Ranson MK, Oliveria-Cruz V, Mills A. Expanding Access to priority health interventions: a framework for understanding to constraints to scaling-up. Journal of International Development 2003; 15: 1-14.

[110] Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Science 2009, 4:50 doi:10.1186/1748-5908-4-50.

[111] ExpandNet. Nine steps for developing a scaling-up strategy. WHO 2010.

[112] NICE. Social value judgements. Principle for the development of NICE guidance. National Institute for Health and Clinical Excellence. Second edition. 2008.

[113] Guyatt GH, Oxman AD, Kunz R, Flack-Ytter Y, Vist GE, Liberati A, Schünemann HJ. GRADE: going from evidence to recommendations. BMJ 2008; 336; 1049-51.

[114] Nixon J, Khan KS, Kleijnen J. Summarising economic evaluations in systematic reviews: a new approach. BMJ 2001;322:1596–8. Quoting: Birch S, Gaffni A. Cost-effectiveness and cost utility analyses: methods for the non-economic evaluation of healthcare programs and how we can do better. In: Geisler E, Heller O, eds. Managing technology in healthcare. Norwell, MA: Kluwer Academic, 1996.

[115] Eddy DM. Comparing benefits and harms: the balance sheet. JAMA 1990, 263:2493. 2498, 2501.

[116] Oxman AD, Lavis JN, Fretheim A, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP) 16: Using research evidence in balancing the pros and cons of policies. Health Research Policy and Systems 2009, 7(Suppl 1):S16. doi:10.1186/1478-4505-7-S1-S16.

[117] Bradick M, Stuart M, Hrachovec J. The use of balance sheets in developing clinical guidelines. JABFP 1999; 12(1): 48-54.

[118] Swiss TPH – LATH – UCT. Harvesting evidence. http://www.harvesting-evidence.org/reviews-and-studies/table-interventions-and-results.html.

[119] Lilford RJ, Pauker SG, Braunholtz DA, Chard J. Decision analysis and the implementation of research findings. BMJ 1998; 317: 405-9.

[120] Sartori AM, de Soárez PC, Novaes HM. Cost-effectiveness of introducing the 10-valent pneumococcal conjugate vaccine into the universal immunisation of infants in Brazil. J Epidemiol Community Health. 2010 Sep 30.

[121] Lomas J, Culyer T, McCutcheon C, McAuley L, Law S. Conceptualising and combining evidence for health systems guidance. Final report. CHSRF 2005.

[122] NICE. A guide for patient/carer groups contributing to a technology appraisal. NHS. National Institute for Health and Clinical Excellence. http://www.nice.org.uk/media/D15/87/TAprocessmanual_patients_carers.pdf.

[123] Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 5. Group processes. Health Research Policy and Systems 2006, 4:17 doi:10.1186/1478-4505-4-17

[124] Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database of Systematic Reviews 2006, Issue 3. Art. No.: CD004563. DOI: 10.1002/14651858.CD004563.pub2.

[125] NICE. Patients and the Public. NHS. National Institute for Health and Clinical Excellence. http://www.nice.org.uk/getinvolved/patientsandpublic/patientandpublichome.jsp.

[126] 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(4): 218-24.

[127] Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. Health Technology Assessment 1998, Vol. 2, Num. 3.

[128] Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. AJPH 1984; 74(9): 979-83.

[129] GLIA. Schiffman RN, Dixon J, Brandt C, O'Conell R, Essaihi A, Michel G, Hisiao A. GuideLine Implementability Appraisal v. 1.0. Yale Center for Medical Informatics. http://gem.med.yale.edu/glia/doc/GLIA_v1.pdf.

[130] WHO. Vaccine introduction guidelines. Adding a vaccine to a national immunisation programme: decision and implementation. WHO – Immunization, Vaccines and Biologicals 2005. WHO/IVB/05.18.

[131] WB. Making services work for poor people. World Development Report 2004. World Bank and Oxford University Press. 2003.

[132] Oxman AD, Bjørndal A, Becerra-Posada F, Gibson M, Block MA, Haines A, Hamid M, Odom CH, Lei H, Levin B, Lipsey MW, Littell JH, Mshinda H, Ongolo-Zogo P, Pang T, Sewankambo N, Songane F, Soydan H, Torgerson C, Weisburd D, Whitworth J, Wibulpolprasert S. A framework for mandatory impact evaluation to ensure well informed public policy decisions. Lancet 2010. 30;375(9712):427-31.

[133] UN. The Millennium Development Goals 2010. UN, New York. 2010.

[134] The Paris Declaration on Aid Effectiveness. 2005.

[135] The Accra Agenda for Action. 2008.

[136] OECD. Glossary of Key Terms in Evaluation and Results Based Management. 2002.

[137] Chalmers I. Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations. Ann Am Acad Pol Soc Sci 2003; 589: 22–40.

[138] AusAID. Monitoring & Evaluation Capacity Building Study. 1997.

[139] Global Fund. M&E manual. 2008. http://www.theglobalfund.org/documents/me/ME_Manual_Module_3_en.pdf.

[140] Vietnam – National M&E Framework for HIV prevention and control programs. 2007. http://www.unaids.org.vn/resource/topic/evaluation/me_framework_jun07_e.pdf.

[141] WHO. Health programme evaluation. Guiding principles for its application in the managerial process for national health development. WHO. Geneva, 1981.

[142] Donabedian A. Evaluating the Quality of Medical Care. The Milbank Memorial Fund Quarterly 1966; 44(3): 166-203.

[143] OECD. Evaluating Development Co-Operation Summary of Key Norms and Standards. 2010.

[144] DAC criteria for Evaluating Development Assistance: http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html. Quoting: The DAC Principles for the Evaluation of Development Assistance, OECD (1991), Glossary of Terms Used in Evaluation, in 'Methods and Procedures in Aid Evaluation', OECD (1986), and the Glossary of Evaluation and Results Based Management (RBM) Terms, OECD (2000).

[145] Boerma T, Abou-Zahr C, Bos E, Hansen P, Addai E, Low-Beer D. Monitoring and evaluation of health systems strengthening WHO, 2010.

[146] Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. International Journal of Epidemiology 1999; 28: 10-18.

[147] Institute of Development Studies. Participatory Monitoring and Evaluation: learning from change. IDS Policy briefing. Issue 12. November 1998.

[148] Moynihan R, Oxman AD, Lavis JN, Paulsen E. Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. Report from Kunnskapssenteret (Norwegian Knowledge Centre for the Health Services) No 1, 2008.

[149] NHMRC. How to put evidence into practice: implementation and dissemination strategies. Handbook series on preparing clinical practice guidelines. Endorsed February 2000. National Health and Medical Research Council. Commonwealth of Australia 2000.

[150] Yuan CT, Nembhard IM, AF Stern, Brush JE, Krumholz HM, Bradley EH. Blueprint for the Dissemination of Evidence-Based Practices in Health Care. The Commonwealth Fund 2010. Pub. 1399. Vol. 86.

[151] WHO. WHO guidelines approved by the Guidelines Review Committee. http://www.who.int/rpc/guidelines/en/index.html.

[152] EVIPNet. http://www.evipnet.org/php/index.php.

[153] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, Dilkes H. Presenting the Results of Cochrane Systematic Reviews to a Consumer Audience: A Qualitative Study. Medical Decision Making 2010. doi:10.1177/0272989X10375853.

[154] Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. Journal of Clinical Epidemiology 2010; 63: 607-19.

[155] Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, Young T, Althabe F, Ciapponi A, Marti SG, Meng Q, Wang J, la Hoz Bradford AM, Kiwanuka SN, Rutebemberwa E, Pariyo GW, Flottorp S, Oxman AD. Evidence summaries tailored to health policy-makers in low- and middle-income countries. Bull World Health Organ. 2011 Jan 1;89(1):54-61. Epub 2010 Nov 24.

[156] Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. Journal of Clinical Epidemiology 2009; 63: 620-6.

[157] NHMRC. How to present evidence for consumers: preparation of consumer publications. Handbook series on preparing clinical practice guidelines. Australia. November 1999.

[158] Lomas J. Beyond the sound of one hand clapping: a discussion document on improving health research dissemination and uptake. University of Sydney 1997.

[159] Canadian Health Services Research Foundation. Communication notes. Reader-friendly writing – 1:3:25. http://www.chsrf.ca/Migrated/PDF/CommunicationNotes/cn-1325_e.pdf.

[160] Creating excellence for better places and communities. Local evidence guide. http://www.creatingexcellence.org.uk/ceimages/CE_LocalEvidenceGuide.pdf.

[161] Rosenbaum Sarah. The Effect of Lay Health Worker Programmes on Mother and Child Health and Tuberculosis. http://www.youtube.com/watch?v=0zHHhpE5Rb0.

[162] Swiss TPH – LATH – UCT. Harvesting evidence. http://www.harvesting-evidence.org/.

[163] Swiss TPH – LATH – UCT. Harvesting evidence. http://www.harvesting-evidence.org/reviews-and-studies/grade-evidence-profiles.html.

[164] Glenton C, Paulsen EJ, Oxman AD. Portals to Wonderland: Health portals lead to confusing information about the effects of health care. BMC Medical Informatics and Decision Making 2005, 5:7 doi:10.1186/1472-6947-5-7.

[165] Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. BMC Health Research Policy and Systems 2006, 4:26 doi:10.1186/1478-4505-4-26.

[166] HON. Health On the Net Foundation. http://www.hon.ch/home1.html.

[167] HON. The HON Code of Conduct for medical and health Web sites (HONcode). http://www.hon.ch/HONcode/Patients/Conduct.html.

[168] Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. The Journal of Continuing Education in the Health Professions 2004; 24: S31-7.

[169] Lavis J, Davies H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. J Health Serv Res Policy 2005. 10. Suppl 1:35-48.

[170] Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. Health Research Policy and Systems 2006, 4:27. doi:10.1186/1478-4505-4-27.

[171] Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, Wensing M, Dijkstra R, Donaldson C. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004;8(6).

[172] Perrier L, Mrklas K, Lavis JN, Straus SE. Interventions Encouraging The Use of Systematic Reviews by Health Policymakers and Managers: A Systematic Review. Implementation Science 2011, 6:43 doi:10.1186/1748-5908-6-43.

[173] Thomas LH, Cullum NA, McColl E, Rousseau N, Soutter J, Steen N. Guidelines in professions allied to medicine. Cochrane Database of Systematic Reviews 1999, Issue 1. Art. No.: CD000349. DOI: 10.1002/14651858.CD000349.

[174] Akbari A, Mayhew A, Al-Alawi MA, Grimshaw J, Winkens R, Glidewell E, Pritchard C, Thomas R, Fraser C. Interventions to improve outpatient referrals from primary care to secondary care. Cochrane Database of Systematic Reviews 2008, Issue 4. Art. No.: CD005471. DOI: 10.1002/14651858.CD005471.pub2.

[175] Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, Robertson N. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. Cochrane Database of Systematic Reviews 2010, Issue 3. Art. No.: CD005470. DOI: 10.1002/14651858.CD005470.pub2.

[176] Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database of Systematic Reviews 2006, Issue 2. Art. No.: CD000259. DOI:

10.1002/14651858.CD000259.pub2.

[177] Conterno LO, Mayhew A, Pereira CR. Interventions aimed at improving professional adherence to guidelines for prevention of device-related infections. Cochrane Database of Systematic Reviews 2007, Issue 2. Art. No.: CD006559. DOI: 10.1002/14651858.CD006559.

[178] Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Family Practice 2000; 17(S1): S11-6. http://fampra.oxfordjournals.org/cgi/reprint/17/suppl_1/S11.

[179] Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, O'Connell R. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. BMC Medical Informatics and Decision Making 2005, 5:23. doi:10.1186/1472-6947-5-23.

[180] GLIA: Guideline Implementation Appraisal. http://gem.med.yale.edu/glia/login.htm.

[181] NICE. NICE implementation programme. NHS. National Institute for Health and Clinical Excellence. http://www.nice.org.uk/usingguidance/niceimplementationprogramme/nice_implementation_programme.jsp.

[182] IHI. Improvement map. Institute for Healthcare improvement. http://www.ihi.org/IHI/Programs/ImprovementMap/.

[183] NICE. Measuring the use of NICE guidance. NHS. National Institute for Health and Clinical Excellence. http://www.nice.org.uk/usingguidance/measuringtheuseofguidance/evaluation_and_review_of_nice_implementation_evidence_ernie.jsp.

[184] Lovejoy TI, Revenson TA, France CR. Reviewing Manuscripts for Peer-Review Journals: A Primer for Novice and Seasoned Reviewers. Annals of Behavioural Medicine 2011. Apr 20. [Epub ahead of print].

[185] BioMed Central. Peer-review strategy. http://www.biomedcentral.com/independent/develop/peerreview.

[186] BMJ. Guidance for peer reviewers. http://resources.bmj.com/bmj/reviewers/peer-reviewers-guidance.

[187] EPOC. Effective Practice and Organisation of Care Group. Guide for Review Authors. 3 - Developing a Protocol. http://epoc.cochrane.org/epoc-resources-review-authors.

[188] Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE: Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Qual Health Care 1999; 11(1):21-28.

[189] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J: Are guidelines following guidelines? the methodological quality of clinical practice guidelines in the peer-reviewed medical literature. JAMA 1999; 281:1900-1905.

[190] AGREE. Advancing the science of practice guidelines. http://www.agreetrust.org/.

[191] YMCI. COGS: The Conference on Guideline Standardisation. http://gem.med.yale.edu/cogs/.

[192] Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM: Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med 2003, 139(6):493-498.

[193] Collaboration AGREE: Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003, 12(1):18-23.

[194] The Cochrane Collaboration. Clinical guidelines. Online resources. http://www.cochrane.org/about-us/evidence-based-health-care/webliography/books/clinguide.

[195] AGREE. Appraisal of guidelines for research and evaluation II. AGREE II instrument. The AGREE next steps consortium. May 2009.

[196] AGREE. Comparison of guideline development programmes. http://www.agreecollaboration.org/comparison/.

[197] WHO. Guidelines Review Committee (GRC). http://www.who.int/rpc/guidelines/review_committee/en/index.html.

[198] Yale Center for Medical Informatics. Guideline Development Model. http://gem.med.yale.edu/default.htm.

[199] Lavis JL, Oxman AD, Lewin S, Fretheim A. Introduction. SUPPORT Tools for evidence-informed health Policymaking (STP). Health Research Policy and Systems 2009, 7(Suppl 1):I1 doi:10.1186/1478-4505-7-S1-I1.

[200] Field MJ, Lohr KN (Eds). Guidelines for clinical practice: from development to use. Institute of Medicine, Washington, D.C: National Academy Press, 1992.

[201] John Snow, Inc./DELIVER in collaboration with the World Health Organization. Guidelines for the Storage of Essential Medicines and Other Health Commodities. 2003. Arlington, Va.: John Snow, Inc./DELIVER, for the U.S. Agency for International Development.

[202] The Mexico statement on health research. Knowledge for better health: strengthening health systems. From the ministerial summit on health research. Mexico city, November 16-20, 2004.

[203] World Health Assembly. Resolution WHA58.34. Ministerial summit on Health Research. WHA 2005. http://www.who.int/rpc/meetings/58th_WHA_resolution.pdf.

[204] The Bamako call to action on research for health. Strengthening research for health, development, and equity. From The Global Ministerial Forum on research for health. Bamako, Mali, November 17-19, 2008. http://www.who.int/rpc/news/BAMAKOCALLTOACTIONFinalNov24.pdf.

[205] WHO. Research policy. Guidelines Review Committee. http://www.who.int/rpc/guidelines/review_committee/en/index.html.

[206] Lavis JN, Ross SE, Hurley JE, et al. Examining the role of health services research in public policymaking. Milbank Q 2002; 80: 125–54.

[207] Centre for Evidence Based Medicine (CEBM). Critical appraisal. Systematic reviews. Critical appraisal sheet. http://www.cebm.net/index.aspx?o=1567.

[208] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004 Jun 19;328(7454):1490.

[209] AHPSR. Alliance for Health Policy and Systems Research. http://www.who.int/alliance-hpsr/en/.

[210] Bandolier. Evidence based thinking about health care. http://www.medicine.ox.ac.uk/bandolier/.

[211] CHSR. Canadian Health Services Research Foundation. http://www.chsrf.ca/Home.aspx.

[212] CHSRF. Mythbusters. http://www.chsrf.ca/PublicationsAndResources/Mythbusters.aspx.

[213] CHSRF. Evidence boost. http://www.chsrf.ca/PublicationsAndResources/PastSeries/EvidenceBoost.aspx

[214] CHSRF. Insight and action. http://www.chsrf.ca/PublicationsAndResources/Mythbusters.aspx.

[215] CHSRF. Promising practices in research use. http://www.chsrf.ca/PublicationsAndResources/PastSeries/PromisingPracticesinResearchUse.aspx

[216] The Cochrane Library. http://www.thecochranelibrary.com/view/0/index.html.

[217] EVIPNet Africa. http://www.who.int/rpc/evipnet/africa/en/index.html.

[218] Wellcome Trust. Health Research Capacity Strengthening in Kenya and Malawi. http://www.wellcome.ac.uk/Funding/International/global-health-research/wtdv026103.htm.

[219] McMaster University. Health Systems Evidence. http://www.mcmasterhealthforum.org/healthsystemsevidence/.

[220] REACH-PI. Regional East African Community Health -Policy Initiative (REACH-PI). http://www.kemri.org/index.php/reach-pi-home.

[221] SUPPORT. Supporting Policy relevant Reviews and Trials. http://www.iecs.org.ar/support/iecs-visor-publicaciones.php.

[222] WHO. Reproductive Health Library. http://apps.who.int/rhl/access/en/index.html.

[223] Zambia forum for Health Research. A Zambia knowledge translation platform. http://www.zamfohr.org/.