

This is the author's final draft post-refereeing article published in The Lancet:

Oxman AD, Lavis JN, Fretheim A.
Use of evidence in WHO recommendations.
Lancet. 2007 Jun 2;369(9576):1883-9.
[http://linkinghub.elsevier.com/retrieve/pii/S0140-6736\(07\)60675-8](http://linkinghub.elsevier.com/retrieve/pii/S0140-6736(07)60675-8)

DRAFT: 24 December 2006

The use of evidence in WHO recommendations

Andrew D Oxman, researcher¹

John N. Lavis, associate professor and Canada Research Chair in Knowledge Transfer and Exchange²

Atle Fretheim, researcher¹

1. Norwegian Knowledge Centre for the Health Services
2. Department of Clinical Epidemiology and Biostatistics and Department of Political Science, and Member of the Centre for Health Economics and Policy Analysis, McMaster University

Address for correspondence:

Dr Andy Oxman
Norwegian Knowledge Centre for the Health Services
P.O. Box 7004, St. Olavs plass
N-0130 Oslo, Norway

Phone: +47 48254924
Fax: +47 23255030
Email: oxman@online.no

Acknowledgements

This study was conducted as part of a project (www.practihc.org) funded by the European Commission's 5th Framework International Collaboration with Developing Countries, Research (Contract ICA4-CT-2001-10019). Lavis receives funding as the Canada Research Chair in Knowledge Transfer and Exchange.

Counts

Abstract	268 words
Text	4,686 words
References	35

Abstract

Background: The World Health Organization's regulations, dating back to 1951, emphasize the role of expert opinion in the development of recommendations. Guidelines for WHO guidelines, approved in 2003, instead emphasize systematic reviews of evidence of effects, processes that allow for the explicit incorporation of other types of information (including values), and evidence-informed dissemination and implementation strategies. We examined the use of evidence, particularly evidence of effects, in recommendations developed by WHO departments.

Methods: We conducted interviews with department directors (or their delegates) at WHO headquarters and reviewed a sample of the recommendation-containing reports that were discussed in the interviews (as well as related background documentation). Two individuals independently conducted a thematic analysis of the interviews and reviewed key features of the recommendation-containing reports (and related background documentation).

Findings: Systematic reviews and concise summaries of findings are rarely used. Processes for developing recommendations typically rely heavily on experts in a particular content area and not on representatives of those who will have to live with the recommendations or on experts in particular methodological areas. Relatively little attention has been given to how to best help member states adapt global recommendations or take account of local needs, conditions, resources, costs, and values. Relatively little attention has also been given to roles and responsibilities related to effective dissemination and implementation strategies and their rigorous evaluation.

Interpretation: Progress in the development, adaptation, dissemination and implementation of recommendations for member states will require leadership, the resources necessary for WHO to undertake these processes in a transparent and defensible way, and close attention to the current and emerging research literature related to these processes.

The use of evidence in WHO recommendations

Each year the World Health Organization (WHO) develops a large number of recommendations aimed at many different target audiences, including the general public, healthcare professionals, managers working in health facilities (e.g., hospitals) or regions (e.g., districts), and public policymakers in member states. These recommendations address a wide range of clinical, public health and health policy topics related to achieving health goals. WHO's regulations, dating back to 1951, emphasize the role of expert opinion in the development of recommendations. In the fifty-five years since these regulations were developed, research has highlighted the limitations of expert opinion. For example, expert opinion can differ both across sub-groups and from the opinions of those who will have to live with the consequences of expert recommendations,¹⁻⁸ and experts have been found to use non-systematic methods when they review research, which frequently result in recommendations that do not reflect systematic summaries of the best available evidence.^{9,10}

Evidence of the effects of alternative policies, programs and services is essential for well-informed decisions. Systematic reviews have several advantages over other approaches to amassing evidence of effects.¹¹⁻¹³ Firstly, systematic reviews reduce the risk of bias in selecting studies and interpreting their results. Secondly, they reduce the risk of being misled by the play of chance in identifying studies for inclusion, or the risk of focusing on a limited subset of relevant evidence. Thirdly, systematic reviews provide a critical appraisal of the available evidence and place individual studies or subgroups of studies in the context of all of the relevant evidence. Finally, they allow others to appraise critically the judgements made in selecting studies and the collection, analysis and interpretation of the results. However, systematic reviews can only be as good as the evidence that they summarize. There may be no evidence. When there is evidence, judgements are still required about the quality of the

evidence and, particularly for public health and health policy topics, its applicability in different contexts.¹²

Evidence of effects needs to be complemented by information about needs, factors that may affect whether effectiveness will be realized in the field, available resources, costs, and the values of those who will be affected by the recommendations. Processes that allow for the explicit incorporation of these other types of information, particularly values, have (like systematic reviews) emerged as central to the development of recommendations.¹⁴⁻¹⁸ Moving from evidence to recommendations requires judgements, particularly judgements about goals and about the balance between the desirable and undesirable consequences of choosing one option over another to achieve these goals.

Evidence-informed dissemination and implementation strategies are increasingly recognized as a core part of the business of developing recommendations. Those charged with developing clinical practice guidelines can draw on a systematic review of randomized controlled trials of guideline dissemination and implementation strategies to inform their efforts.^{19,20} While there are no magic bullets and few strategies have been evaluated in low and middle-income countries, such efforts clearly can have an effect.²¹ Those charged with developing recommendations targeted at managers or public policymakers, on the other hand, have to deduce the attributes of the interventions from systematic reviews of observational studies and begin to build an evidence base about the effectiveness of these interventions.^{13,,22}

WHO has recognised the need to revise its approach to developing recommendations, as reflected in the guidelines for WHO guidelines, which were approved by the WHO Cabinet in 2003.²³ We sought to examine the use of evidence in WHO recommendations subsequent to

this. We particularly wanted to explore the use of evidence of effects. Our hope was that such stock-taking would inform debates about how WHO could improve how it develops and disseminates recommendations and how WHO could better support member states in their efforts to adapt and implement recommendations.

Methods

We conducted interviews with department directors (or their delegates) at WHO headquarters and reviewed a sample of the recommendation-containing reports that were discussed in the interviews. We invited the participation of all department directors in five departmental clusters that had a content focus: non-communicable diseases and mental health (six departments); HIV/AIDS, TB and malaria (four); family and community health (four); communicable diseases (three); and health technology and pharmaceuticals (two). We invited the participation of one department (of five) in the sustainable development and healthy environments cluster and three departments (of five) in the evidence and information for policy cluster. We did not invite the participation of the department directors in the two clusters -- the external relations and governing bodies cluster and the general management cluster -- that had a corporate focus. While our e-mailed request for an interview was introduced by a WHO department director, we made clear that we were conducting the study independently of WHO and that we planned to publish the results after first making them available to WHO. We purposively sampled four of the reports identified by the interviewees on the basis of their focus on clinical treatment, centrality to major WHO initiatives, and relevance to the Millennium Development Goals. The first criterion was chosen to maximize the chances that evidence of effects would be available and that the reports could be expected to meet current standards for clinical practice guidelines.

Two individuals participated in each interview. One individual had primary responsibility for conducting the interview and the other for audio taping the interview and taking notes. The brief structured part of each interview focused on the number and background of staff members and the number and type of recommendation-containing reports published in the last year. The semi-structured part of each interview focused primarily on the development of recommendations contained in one or two specific published guidelines or policies that were selected by interviewees from among those their department had developed or had a major responsibility in developing. For each guideline or policy we asked about: why it was developed; the process used (including whether support was received from others within or outside WHO, whether evidence of effects and other types of information were used, whether and how supporting documentation was made publicly available, and whether and how plans for updating were established); strengths of the processes used and elements that could have been improved upon; likely benefits, harms and costs of adhering to the recommendations; how the recommendations have been used and any plans for evaluating the effects of adherence to the recommendations; and the availability of any background documentation.

Two individuals independently conducted a thematic analysis of the interviews and reviewed key features of the recommendation-containing reports (and related background documentation). We began the thematic analysis by using the notes taken during each interview (supplemented by the corresponding audiotape) to produce a summary of each interview, including the major themes that emerged. We then sent the summary to each interviewee with a request that they verify our interpretations and, if they wished, provide additional comments and/or reflections. We used the audiotapes to identify illustrative quotations for each major theme. We began the document review by recording for each document its type, whether it included a section that described the methods that were used,

the number of recommendations that were based on a systematic review, the number of systematic reviews cited, and the description that was provided of the development process that was used. We then produced a summary for each recommendation-containing report. We presented our findings at a variety of forums within WHO as an additional check on our interpretations.

The study was sponsored as part of a broader project -- Pragmatic randomized controlled trials in health care (www.practihc.org) -- funded by the European Commission's 5th Framework International Collaboration with Developing Countries, Research. WHO was a formal partner in Practihc, however, as a partner WHO staff input was limited to commenting on the protocol and interview questions. Additional input from WHO staff was limited to providing comments on the interview summaries (for those who were interviewed) and overall findings (for those who attended forums where these were presented). The sponsors of the study had no role in the design (beyond commenting on the protocol and interview questions), data collection, data analysis, data interpretation (beyond commenting on the interview summaries and overall findings), writing or revising of the report. ADO, as corresponding author, had full access to all data in the study and had final responsibility to submit a report for publication.

Results

We conducted 23 interviews with 29 people and we reviewed four recommendation-containing reports and related background documentation. We interviewed the director in 15 departments and someone designated by the director in six departments. Three of the interviews were with more than one person. For two departments we conducted two separate interviews with different people from the department. We were unable to arrange interviews with the directors (or delegates) of two departments that had newly appointed directors who

were not yet in post. The interviews, which lasted one hour or less, were conducted between September 2003 and February 2004. Five of the 21 participating departments did not produce formal recommendations so their interviews were not included in the analysis, which is therefore based on 17 interviews with 21 people (across 16 departments). The four recommendation-containing reports that we selected for review were clinical practice guidelines that addressed antiretroviral therapy for HIV, treatment of tuberculosis, treatment of malaria, and Integrated Management of Childhood Illness (IMCI).²⁴⁻²⁷

Who is producing what types of recommendations and why?

The directors or their delegates (hereafter directors) of the 16 departments that developed recommendations reported that their departments had between eight and 170 staff members each (median 55) and close to 1000 staff members in total. The directors estimated that between 20% and 80% of staff members had some background in research (median 30%). Many directors found it difficult to quantify the number of recommendation-containing reports that their department published each year given the variety of formats in which recommendations could appear. Their estimates ranged from one to 45 reports per department per year (median 8) and totalled close to 180 reports per year.

The reports varied widely in the nature of the topics they addressed. In addition to clinical treatment topics, the reports addressed topics like malaria control using impregnated bed nets, promotion of mental health, helminthic guidelines for managers, human resources policy development, model list of essential medicines, tobacco legislation, and bioterrorism. The directors cited a number of reasons for developing recommendations, the most common of which were a perceived need for guidance, a perceived need for updating existing

recommendations, and demand from member states. One report was developed to respond to criticisms of previous recommendations.

How are recommendations being developed?

Expert committees or meetings of experts were almost always convened when developing recommendations whereas only a few directors mentioned having commissioned systematic reviews to inform the work of these expert groups. Some directors reported the use of a combination of work done “in house” and an expert committee or the combination of a small “task force” to draft recommendations and either an expert committee or a review by external experts. A fair number of directors reported a phase of external consultation or review. Only a few directors mentioned developing dissemination or implementation strategies. Most directors reported the involvement of one or more other WHO departments in the development process and nearly all reported some form of external support. No directors mentioned drawing on any form of internal support in the methodological or technical aspects of developing recommendations. The external support typically took the form of expert committee members, but sometimes involved expert advisors, writers of background reports and recommendation-containing reports, and reviewers.

When asked specifically about using evidence of effects, only a small number of directors reported using systematic reviews of such evidence and none reported using concise summaries of findings (e.g., balance sheets) for the most important outcomes (benefits, harms and costs) of each option being considered. Many directors instead reported using background documents, although there was little consistency in how the documents were prepared. For example, some background documents were prepared by the participating experts according to their own conventions. Other directors reported leaving the use of evidence up to the

experts, feeling that evidence of effects was not relevant for some recommendations, and feeling that randomized trials were not appropriate for some types of interventions. Only one director reported grading the quality of the evidence.

When asked about using other types of information, a number of directors reporting using data about costs but only a couple mentioned using data about potential harms or explicitly considering values; i.e. the relative importance or worth of the consequences (benefits, harms and costs) of a decision.¹⁸ Using data about potential harms was only mentioned in relation to clinical interventions, particularly pharmaceuticals, and not for public health or policy interventions. Explicitly considering values was undertaken in a very general way. One director talked about the “weighing of values, which basically reflected the composition of the panel.” Another director commented:

“Values were also brought into debate. For example, experience for high income countries suggest that encouraging more self efficacy and independency for young people could be effective in preventing mental health problems and substance use. However, this was by many considered as to be contrary to important values for people living in many low-income countries.”

While directors were not asked specifically about group processes, many volunteered descriptions that suggested that these processes were not particularly structured with respect to group composition, format or rules. The directors’ descriptions suggested that participants were implicitly weighing evidence of effects, harms, and costs along with values and many other types of information (e.g., surveys, resistance patterns, other epidemiological data, availability of interventions, country experiences, political considerations, cultural

differences, ethical considerations, and “undocumented knowledge”). One director clearly recognized the challenges associated with a lack of structured process:

“There is a tendency to get people around the table and get consensus (everything they do has a scientific part and a political part). This usually means you go to the lowest common denominator or the views of a “strong” person at the table.”

Most directors reported that the information that was used by the committees was not published but that it was often made publicly available in some form. The format for the documentation varied widely, including a bibliography in the report, one or more published articles (e.g., a special edition of the WHO Bulletin), one or more reports (e.g., annual reports, multi-country evaluation reports, and proceedings of meetings), a book, and an adaptation guide. The documentation was sometimes readily available (e.g., on a website) and other times required personal contact with those involved in developing the recommendations. While one director reported updating recommendations every two years and some other directors reported that their recommendations were considered one-off initiatives that would not be updated, most directors reported ad hoc approaches to deciding whether and when to update recommendations. One director reported plans to update the recommendations using the guidelines for WHO guidelines.

What is being done well and what could be improved?

The directors identified a number of strengths in the processes used for developing recommendations and in the recommendations themselves. The most commonly identified strength was bringing together or consulting with a wide range of people. Most other

strengths were mentioned by only one director, although several of these strengths can be grouped together:

- usefulness of the recommendations, which included attributes like focusing on end users, ensuring usability, responding to the concerns of donors, and filling a gap;
- evidence-based process, which included attributes like obtaining evidence in a rigorous way, drawing on good data, basing recommendations on research, using cost-effectiveness analyses, testing the recommendations, and conducting validation studies;
- experience-based process, which included attributes like involving people with practical experience and, while this was also considered a weakness, developing “instinct-based” recommendations;
- expert-based process, which included attributes like working with knowledgeable experts and obtaining consensus among experts;
- systematic approach, which included attributes like using a standardised method and adopting “guideline logic” rather than “technocratic” approach;
- group members without conflicts of interest;
- good group process as a key element of the meeting structure; and
- up-to-date recommendations.

While most directors identified one or more ways in which the recommendation-development process could have been improved, four did not identify any way in which improvements could have been made. Directors singled out the use of evidence more commonly than any other area for improvement. Comments included:

- “I would have liked to have had more evidence to base recommendations on. We should have conducted a literature search.”

- “We never had the evidence base well documented. We should have reviewed evidence at a very early stage.”
- “The lack of resources does limit the ability to develop evidence-based recommendations.”
- “[Director General] Brundtland came in and said “evidence, evidence, evidence” but the approach to expert committees hasn’t changed since the 1950s – many see WHO as a technical agency and therefore we should have a comprehensive review of recommendation processes, including expert committees.”
- “Maybe what WHO needs is more work on the guidelines for guidelines.”

Directors also frequently singled out the timeliness of the recommendations as an area for improvement. Directors offered comments like “It could have happened earlier” and “It could have been done faster... perhaps better with one person being responsible for keeping up the momentum.” Recommendations were sometimes prepared as a “technical consultation” document as a way of reducing both the amount of time needed to produce recommendations and the level of expectations about the rigour of the process used. One director described a recommendation-containing report that was:

“... prepared as a technical consultation document so it has a lower status. They should have been prepared by a study group and, even better, an expert committee. People have asked how can you say a technical consultation document is a WHO recommendation, but it has stood the test of time with other initiatives coming to similar conclusions. You can’t develop a guideline in less than a year, but this doesn’t work when there’s pressure. Should there be a guideline for urgent recommendations?”

Several directors identified the match between the resources available and the resources needed to develop recommendations and attention to dissemination and implementation strategies as other areas for improvement. Two directors identified a lack of resources as the problem. For example, one said: “We had inadequate time and resources. The recommendation was developed during about ten months. I believe this is too short a time. Would like to be able to use a more systematic approach.” Two other directors indicated that the resources required to develop recommendations was the problem. For example, one said: “It was a cumbersome and resource-demanding exercise.” Several directors observed that recommendations were not being implemented after they were published. One said: “We published it, but just left it there . . . The recommendations were never transformed into a programmatic approach. It is a common in-house failure to transform recommendations into action.” Another said: “The marketing of it, making people aware, should have been thought of earlier.”

Directors highlighted a number of other weaknesses with the processes used to develop recommendations, however, most were mentioned by only one director. The weaknesses included a failure to involve key organizations, a failure to use evidence from other sectors, the creation of high expectations, a conflict over data, failure to use the guidelines for WHO guidelines (which were published after the process was started), the perceived need to choose between having a “mega-meeting” or using a smaller group to develop recommendations, the failure to involve patients sufficiently, the failure to fit recommendations to health systems, not having had consultations earlier in the process, and not collecting baseline data for an evaluation.

What is expected and used and with what effect?

The anticipated benefits, potential harms and costs of adherence to the recommendations were unevenly considered. All directors could cite one or more anticipated benefits of adherence, such as simplification of treatment, improved quality of care, better management of technologies, and reduced morbidity and mortality. Many fewer directors could cite one or more potential harms of adherence. Indeed, a number of directors reported that there were no potential harms in adhering to their departments' recommendations. For example, one director argued: "No harms are likely, since the recommendations were made by the top experts." Those directors who could cite potential harms provided relatively general examples, such as side effects and the consequences of misapplication or adaptation of the recommendations. One director reported that the potential harms were only considered implicitly in the discussion because it was feared that emphasizing the risks might reduce the value of the recommendations, which were intended to help countries advocate for disease control programs. Many directors identified both direct costs and opportunity costs associated with adhering to the recommendations.

When asked about how their department's recommendations have been used, directors provided examples like educators using them in training programs, WHO staff using them in their work in countries, and member states using them in developing policies. Several directors reported requests for reports, webpage hits, or translations of reports as indicators of the usefulness of the recommendations. Only a few directors reported any systematic monitoring of the uptake of their recommendations. Similarly, only a few reported completed or planned evaluations. This may be due, in part, to a lack of resources, as suggested by one director: "We would love to do it through a rigorous process. The problem is that that this would require resources that we do not currently have and cannot reasonably expect in the foreseeable future." Another reported "no plans for evaluation because the cycle of scientific

developments is so quick that it isn't feasible." Most of the reported evaluations were not rigorous evaluations of their effectiveness. They included the collection of indicators, case studies, site visits, and feedback at meetings.

How were four clinical guidelines developed?

The four clinical guidelines that we examined did not emphasize evidence about effectiveness or processes that would allow for the explicit incorporation of other types of information.²⁴⁻²⁷ Two of the reports were called "guidelines", one a "technical consultation", and another a description of "the technical basis for the guidelines". Two of the reports stated that "This document is not a formal publication of WHO" on the page containing the publication information. Three of the reports (including the supporting documentation) did not contain a methods section. The fourth report contained a brief (less than one page) methods section. In all four reports the recommendations were neither itemized nor explicitly linked to evidence. All reports included references to primary studies or secondary sources. Three of the reports cited at least one systematic review as a reference (and at most four). The descriptions of the recommendation-development processes used in developing the recommendations were brief and provided little information about group processes (Box 1).

What if any progress is being made?

A number of directors indicated that there was a growing recognition of the need for more systematic and transparent approaches to developing recommendations and that there was progress in this direction. One director observed: "There has been a culture change, but there is room for improvement." Another director said: "It is improving, but slowly. Many departments are doing OK, while others are not doing so well. Some have been too close to industry, often because of lack of resources." A third director also provided a long-range

view: “We are in the middle of a process, which needs time. There is increasing understanding of the need for evidence-based guidance and it is becoming part of the WHO culture.”

Discussion

The guidelines for WHO guidelines do not appear to be closely followed when WHO develops recommendations for member states. For example, systematic reviews and concise summaries of findings (e.g., balance sheets) are rarely used, which means that evidence is generally not retrieved, appraised, synthesized and interpreted using systematic and transparent methods. Processes for developing recommendations typically rely heavily on experts in a particular content area and not on representatives of those who will have to live with the recommendations or on experts in particular methodological areas (e.g., information retrieval, systematic reviews, economic evaluations, and group facilitation). While many of the people we spoke with viewed this as a problem, many others did not. Relatively little attention appears to have been given by WHO to how to best help member states adapt global recommendations or take account of local needs, conditions, resources, costs and values. Relatively little attention has also been given by WHO to roles and responsibilities related to effective dissemination and implementation strategies and their rigorous evaluation.

The strengths of our study include achieving a high response rate among the directors of a broad cross-section of WHO departments, conducting interviews that probed the contexts for and processes used in developing specific guidelines or policies, augmenting the interviews with document reviews in a domain that could be expected to be a best-case scenario (developing clinical practice guidelines as opposed to public health or policy recommendations), and undertaking two efforts to verify our interpretations (sharing our

written summaries of each interview with directors and sharing our findings at a variety of forums within WHO). The verification process yielded only minor corrections. The study's weaknesses include the potential for social desirability bias, particularly in terms of identifying the use of evidence as an area for improvement.

While the guidelines for WHO guidelines are consistent with guidelines developed by other organizations,²⁸ the actual processes used to develop recommendations at WHO appear to be less rigorous than at other organizations. None of the directors reported using the guidelines for WHO guidelines and only two reported plans to use them. Few directors reported using processes that were consistent with the guidelines. An unpublished in-house review of WHO guidelines, which was undertaken just prior to our study using the AGREE instrument,¹⁴ found that the vast majority of WHO guidelines did not meet most of the AGREE criteria (Personal communication, Robin Gray, 8 September 2003). Reviews of clinical practice guidelines produced by other organizations have also found that guidelines often do not adhere to guidelines for guidelines.²⁹⁻³¹ WHO also is not alone in its failure to recognise the danger of inadequately evaluated public health and policy interventions which, like clinical interventions, can also have unintended consequences.³² However, many organizations now report using systematic and transparent methods to develop clinical, public health and policy recommendations, including a growing number of organizations funded by government.³³⁻³⁵

Progress in how WHO develops and disseminates recommendations for member states and in how it supports member states in their efforts to adapt and implement recommendations will require leadership. WHO's Cabinet recognised the need for using systematic and transparent methods to develop recommendations when it endorsed the

guidelines for WHO guidelines in 2003.²³ Yet no mechanisms have been put in place to support and monitor adherence to the guidelines and our study suggests that the guidelines are not being followed. A number of directors reported a shift towards a culture that supports using systematic and transparent methods in developing recommendations but this shift appeared to pertain more to clinical than policy recommendations. WHO has not clearly articulated whether and how it will support member states in their efforts to adapt and implement recommendations.

Progress will also require the resources necessary for WHO to undertake recommendation-development processes in a transparent and defensible way and close attention to the current and emerging research literature related to these processes. All of the directors we interviewed were highly motivated and trying hard to do a good job. Many were frustrated by a lack of resources and feelings of being pressured by a lack of time and perceptions of urgency. WHO relies heavily on external financial support, so it could be argued that resources will have to be found outside the organization. However, WHO could do much better with the resources it has, both by setting priorities for guidelines and by adhering to its own guidelines for WHO guidelines. Given that these guidelines may be most germane to the development of clinical practice guidelines and public health recommendations, future iterations of the guidelines will need to incorporate the emerging research literature about developing policy recommendations.¹²

Contributors

All of the authors contributed to the design of the study, participated in interviews, contributed to the analysis and drafting of the report, and have read and approved the final version of the report.

Conflicts of interest

AO is a member of the WHO Advisory Committee on Health Research.

Acknowledgements

We are grateful to all of the people who agreed to be interviewed and thank them for taking time from their busy schedules to meet with us. We would also like to thank colleagues at WHO for their advice and comments on earlier versions of this report.

References

1. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; 2:i-88.
2. Herrin J, Etchason JA, Kahan JP, Brook RH, Ballard DJ. Effect of panel composition on physician ratings of appropriateness of abdominal aortic aneurysm surgery: elucidating differences between multispecialty panel results and specialty society recommendations. *Health Policy* 1997; 42:67-81.
3. Ayanian JZ, Landrum MB, Normand SL, Guadagnoli E, McNeil BJ. Rating the appropriateness of coronary angiography--do practicing physicians agree with an expert panel and with each other? *N Engl J Med* 1998; 338:1896-1904.
4. Fitch K, Lazaro P, Aguilar MD, Martin Y, Bernstein SJ. Physician recommendations for coronary revascularization. Variations by clinical speciality. *Eur J Public Health* 1999; 9:181-7.
5. Vader JP, Porchet F, Larequi-Lauber T, Dubois RW, Burnand B. Appropriateness of surgery for sciatica: reliability of guidelines from expert panels. *Spine* 2000; 25:1831-6.
6. Bernstein SJ, Lazaro P, Fitch K, Aguilar MD, Kahan JP. Effect of specialty and nationality on panel judgments of the appropriateness of coronary revascularization: a pilot study. *Med Care* 2001; 39:513-20.
7. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ* 2001; 323:1218-22.
8. Raine R, Sanderson C, Hutchings A, Carter S, Larkin K et al. An experimental study of determinants of group judgments in clinical guideline development. *Lancet* 2004; 364:429-37.

9. Oxman AD, Guyatt GH. The science of reviewing research. *Annals of the New York Academy of Science* 1993; 703:125-34.
10. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC: A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Treatments for myocardial infarction. JAMA* 1992; 268:240-8.
11. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; 309:597-9.
12. Lavis JN, Posada FB, Haines A, Osei E. Use of research to inform public policymaking. *Lancet* 2004; 364:1615-21.
13. Lavis JN, Davies HTO, Oxman AD, Denis J-L, 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):S35-48.
14. Hayward RS WM, Tunis SR, Bass EB, Rubin HR, Haynes RB: More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 1993; 118:731-7.
15. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12:18-23.
16. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ* 2004; 329:224-7.
17. Culyer AJ, Lomas J. Deliberative processes and evidence-informed decision-making in health care: Do they work and how might we know? *Evidence & Policy* 2006; 2:357-71.
18. Schünemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. *Health Res Policy Syst* 2006, 4:22 (5 December 2006).

19. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001; 39:II2-45.
20. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004; (8)1-72.
21. Haines A, Kuruville S, Borchert M. Bridging the implementation gap between knowledge and action for health. *Bull World Health Organ* 2004; 82:724-33.
22. Innvaer S, Vist GE, Trommald M, Oxman AD. Health policy-makers' perceptions of their use of evidence: a systematic review. *J Health Serv Res Policy* 2002; 7:239-44.
23. Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines. Geneva: World Health Organization, 2003 (EIP/GPE/EQC/2003.1).
24. WHO. Scaling up antiretroviral therapy in resource-limited settings: Guidelines for a public health approach. Geneva: World Health Organization, 2002.
25. WHO. Treatment of tuberculosis: Guidelines for national programmes 3rd edition. Geneva: World Health Organization, 2003.
26. WHO. Antimalarial drug combination therapy: Report of a WHO technical consultation. Geneva: World Health Organization, 2001.
27. Gove S. Integrated management of childhood illness by outpatient health workers: technical basis and overview. *Bull World Health Organ* 1997; 75: 7-24.
28. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Res Policy Syst* 2006, 4:13 (21 November 2006).
29. 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; 281:1900-5.

30. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000; 355:103-6.
31. Burgers JS, Grol R, Klazinga NS, Mäkelä M, Zaat J for the AGREE Collaboration. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Quality Health Care* 2003; 15:31-45.
32. Chalmers I. Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations. *Annals of the American Academy of Political and Social Science* 2003; 589:22-40.
33. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J; AGREE Collaboration. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care* 2003; 15:31-45.
34. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder H. International comparison of the definition and the practical application of health technology Assessment. *IntJ Technol Assess Health Care* 2005; 21:89–95.
35. Moynihan R, Oxman AD, Lavis JN, Paulsen E. Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. A review of organizations that support the use of research evidence in developing guidelines, technology assessments, and health policy, for the WHO Advisory Committee on Health Research. Oslo: Norwegian Knowledge Centre for the Health Services, 2006.

Box 1: Recommendation-development processes used in four guidelines

The following descriptions are taken from the four guidelines included in our document review.

1. “. . . year-long process of international consultative meetings in 2001, in which more than 200 clinicians, scientists, government representatives, representatives of civil society and people living with . . . from more than 60 countries participated.”

“The recommendations included in this publication reflect the best current practices based on a review of existing evidence. When the body of evidence was not conclusive, expert consensus was used as a basis for recommendations.”

2. “This document was prepared for the WHO . . . by . . . The document was reviewed by the WHO Regional Advisors . . . and approved by the WHO Strategy and Technical Advisory Group . . .”

3. “A WHO Technical Consultation on . . . was held in Geneva, Switzerland on 4 and 5 April 2001. Participants reflected a wide range of expertise in the document and use of . . . drugs.

“The technical consultation took the form of presentations based on working papers and plenary discussions, on the basis of which specific conclusions and recommendations were agreed. The proceedings of the meeting and working papers form the basis of this report.”

4. “The guidelines . . . are based on both expert clinical opinion and research results. A technical review of existing programme guidelines was carried out with the cooperation of 12 WHO technical programmes through the WHO Working Group on . . . Some modifications were required . . . The draft guidelines were subsequently reviewed in several versions by clinicians and experts in specific diseases who had experience in clinical and public health work in developing countries, then examined in research studies and by field-testing the training course.”

“Sufficient data were not available to make several guideline decisions. . . . Six studies were carried out . . .”

“The case management charts and the modules were revised based on this experience and on the results of additional studies and analyses to help identify the best clinical indicators . . .”

“The revised materials were made available to countries for closely monitored use . . .”