

Improving the Use of Research Evidence in Guideline Development



Improving the Use of Research Evidence in Guideline Development



Washington DC
2010

Published also in Spanish with the title:
*Mejoramiento del uso de los datos de investigación
para la formulación de directrices*

© 2006 Schünemann et al. Licensee BioMed Central Ltd.

Research publication series

This publication is a reproduction of the collection of articles originally published in the journal *Health Research Policy and Systems*. It was also reproduced in the section "Online Extras" of the *Pan American Journal of Public Health*. The collection is offered in print format to facilitate its access by those who use it as reference material but have limited access to the Internet.

This is an open access collection of articles distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The full collection is available in electronic format at <http://www.health-policy-systems.com>. and <http://journal.paho.org>

You may also access other research publications at www.paho.org/researchportal

Keywords

Guidelines, Guideline development, scientific evidence.

A product of the Research Promotion and Development team, HSS, of PAHO with the permission of Biomed Central.

Printing was financed with funds provided by the Agencia Española de Cooperación Internacional para el Desarrollo, AECID, to PAHO

Contents

Introduction	4
1. Guidelines for guidelines	7
2. Priority setting.....	13
3. Group composition and consultation process	20
4. Managing conflicts of interests	26
5. Group processes	32
6. Determining which outcomes are important.....	36
7. Deciding what evidence to include	42
8. Synthesis and presentation of evidence.....	49
9. Grading evidence and recommendations	59
10. Integrating values and consumer involvement.....	66
11. Incorporating considerations of cost effectiveness, affordability and resource implications.....	74
12. Incorporating considerations of equity	80
13. Applicability, transferability and adaptation	86
14. Reporting guidelines	96
15. Disseminating and implementing guidelines	102
16. Evaluation	106

Review

Improving the use of research evidence in guideline development: introduction

Andrew D Oxman*¹, Atle Fretheim¹, Holger J Schünemann² and SURE³

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway, ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ³Subcommittee on the Use of Research Evidence (SURE) of the WHO Advisory Committee on Health Research (ACHR)

Email: Andrew D Oxman* - oxman@online.no; Atle Fretheim - atle.fretheim@nokc.no; Holger J Schünemann - hjs@buffalo.edu

* Corresponding author

Published: 20 November 2006

Received: 30 March 2006

Health Research Policy and Systems 2006, 4:12 doi:10.1186/1478-4505-4-12

Accepted: 20 November 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/12>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In 2005 the World Health Organisation (WHO) asked its Advisory Committee on Health Research (ACHR) for advice on ways in which WHO can improve the use of research evidence in the development of recommendations, including guidelines and policies. The ACHR established the Subcommittee on the Use of Research Evidence (SURE) to collect background documentation and consult widely among WHO staff, international experts and end users of WHO recommendations to inform its advice to WHO. We have prepared a series of reviews of methods that are used in the development of guidelines as part of this background documentation. We describe here the background and methods of these reviews, which are being published in *Health Research Policy and Systems* together with this introduction.

Background

In May of 2005 the 58th World Health Assembly passed a resolution requesting the Director-General "to undertake an assessment of WHO's internal resources, expertise and activities in the area of health research, with a view to developing a position paper on WHO's role and responsibilities in the area of health research, and to report through the Executive Board to the next World Health Assembly." Related to these resolutions, WHO has asked the Advisory Committee on Health Research (ACHR) for advice on ways in which WHO can improve the use of research evidence in the development of recommendations, guidelines and policies.

The ACHR established a subcommittee to collect background documentation and consult widely among WHO staff, international experts and end users of WHO recommendations to inform this advice. The advice will focus

on processes to ensure that WHO's recommendations are well informed by the best available research evidence. These processes range from how WHO sets priorities for the development of recommendations to how its recommendations are disseminated and implemented, including recommendations developed at WHO headquarters in Geneva, at its regional offices and in countries.

WHO from its inception has focused on research, which is mandated in its constitution, and has been a leading player in the global effort to strengthen ties between research and health development.

Given WHO's position as the world's leading public health agency, it is essential that the organisation, its leaders and its governing body ensure that its recommendations and actions are as well informed as possible by the

best available research evidence. WHO has strived to do this for over 50 years with much success.

However, major new developments have occurred since WHO was established that have led governments around the world to reconsider the methods that they use to ensure that decisions about health care and public health are well informed by research evidence. This reflection and subsequent changes in how recommendations about health are developed have been driven by recognition of gaps between available research evidence and what is done in practice, variations in practice and outcomes, concerns about the quality of health care, and rising health care costs.

Increasingly governments, professional and consumer organisations are demanding more rigorous processes to ensure that health decisions are well informed by the best available research evidence. The processes, in contrast with traditional approaches that rely heavily on the opinions of experts, demand systematic and transparent approaches to access, synthesise and interpret research evidence; and to integrate that evidence with the other information, values and judgements to formulate recommendations. The need for more rigorous processes is underscored by evidence of inconsistencies between the available evidence and expert recommendations [1,2], insufficient use of the available evidence [3,4], and other insufficiencies in how guidelines and recommendations are developed [5-12]. Similar criticisms have been raised and calls have been made for better use of research evidence for health care management and policy making, as well [13-15].

WHO has the opportunity and the mandate to capitalise on these advances and to assist its member states to do so. This is essential to ensure that decisions about health are well informed by research evidence, and that these decisions lead to effective, efficient and equitable actions towards achieving WHO's goal: the attainment by all peoples of the highest possible level of health.

As part of the background documentation to inform ACHR's advice to WHO we have prepared a series of reviews on the following topics:

- Guidelines for guidelines [16]
- Priority setting [17]
- Group composition and consultation process [18]
- Managing conflicts of interest [19]
- Group processes [20]

- Determining which outcomes are important [21]
- Deciding what evidence to include [22]
- Synthesis and presentation of evidence [23]
- Grading evidence and recommendations [24]
- Integrating values and consumer involvement [25]
- Incorporating considerations of cost-effectiveness, affordability and resource implications [26]
- Incorporating considerations of equity [27]
- Adaptation, applicability and transferability [28]
- Reporting guidelines [29]
- Disseminating and implementing guidelines [30]
- Evaluation [31]

We have used the term 'guidelines' broadly to include a wide range of recommendations that WHO makes, including clinical, public health and health policy recommendations. Although much of the literature that we have reviewed has focused on clinical practice guidelines, we have tried to incorporate corresponding literature for public health guidelines and health policy recommendations.

The reviews are not full systematic reviews, although we have aimed to be reasonably systematic and transparent about the methods we have used and the basis for the recommendations that we have made. For each review we began with a series of key questions that were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). The first author of each review conducted searches for relevant literature and prepared the first draft. The search strategies that were used are summarised in each review. We did not always conduct exhaustive reviews. We tried first to identify existing systematic reviews that addressed the questions that we asked and, secondarily, if we did not find a systematic review, relevant methodological research. When there was a paucity of research, we have also included some descriptive literature or, in some cases, evidence that was not directly related to guidelines development.

Each review includes short summaries of what WHO and other organisations are doing, our key findings in relationship to each of the questions that we asked, a discussion of those findings and some suggestions for further work that is needed. Our answers to the key questions that we asked are summarised in the abstract of each review.

A draft of each review was first discussed and revised by the authors. The reviews were also circulated to the ACHR SURE members and discussed by the subcommittee. After peer review the articles were revised by the authors and updated if necessary. We are grateful to the editors of *Health Research Policy and Systems* for agreeing to publish these papers in their journal. In addition to benefiting from their editorial support, this has enabled us to take advantage of the BioMed Central's open peer review system to help ensure the quality of our reviews and advice. We also believe that these reviews are of wide interest to other organisations and individuals that are responsible for developing guidelines or health policy.

In addition to this series of reviews that is being published in *Health Research Policy and Systems*, we have conducted reviews of what WHO is currently doing, using both document analyses and interviews and we have conducted a survey of initiatives around the world that support the use of research evidence in developing guidelines or health policy. We have referred to these reports, which are being published separately, where relevant in the reviews in this series.

Preliminary advice from the ACHR has already been discussed with the leadership of WHO. It has been positively received and, to some extent, is being acted upon already. Before delivering our final report and advice to WHO, we will consult with a reference panel and others within and outside of WHO. We look forward to working with WHO to help implement this advice. We hope that it will assist WHO to better serve its member states by ensuring that its recommendations are well-informed by the best available research evidence, and by enabling those responsible for making decisions to make well-informed choices.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this introduction. AF and HJS contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these

reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series. The members of SURE include: Andy Haines, Ana Langer, Andy Oxman (chair), Tikki Pang (WHO), Ulysses Panisset (WHO), Judith Whitworth (chair of the ACHR).

References

- Oxman AD, Guyatt GH: **The science of reviewing research.** *Annals of the New York Academy of Science* 1993, **703**:125-34.
- 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**:240-248.
- Silagy CA, Stead LF, Lancaster T: **Use of systematic reviews in clinical practice guidelines: case study of smoking cessation.** *BMJ* 2001, **323**:833-36.
- Vigna-Taglianti F, Vineis P, Liberati A, Faggiano F: **Quality of systematic reviews used in guidelines for oncology practice.** *Annals of Oncology* 2006, **17**:691-701.
- Bradbury J: **Storm over WHO-ISH hypertension guidelines.** *Lancet* 1999, **353**:563.
- Horton R: **WHO: the casualties and compromises of renewal.** *Lancet* 2002, **359**:1605-11.
- Laing R, Waning B, Gray A, Ford N, t Hoen E: **25 years of the WHO essential medicines lists: progress and challenges.** *Lancet* 2003, **36**:1723-9.
- McCarthy M: **Critics slam draft WHO report on homoeopathy.** *Lancet* 2005, **366**:705.
- Oxman AD, Lavis JN, Fretheim A: **The use of evidence in WHO recommendations.** *Lancet* in press.
- 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.
- 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.
- Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H: **Attributes of clinical guidelines that influence use of guidelines in general practice: observational study.** *BMJ* 1998, **317**:858-61.
- Lavis JN, Posada FB, Haines A, Osei E: **Use of research to inform policymaking.** *Lancet* 2004, **364**:1615-21.
- Lavis JN, Davies H, Oxman AD, 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.
- Sheldon TA: **Making evidence synthesis more useful for management and policy-making.** *J Health Serv Res Policy* 2005, **10**(Suppl 1):1-5.
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 1. Guidelines for guidelines.** *Health Res Policy Syst* 2006.
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 2. Priority setting.** *Health Res Policy Syst* 2006.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 3. Group composition.** *Health Res Policy Syst* 2006.
- Boyd E, Bero L: **Improving the Use of Research Evidence in Guideline Development: 4. Managing conflicts of interest.** *Health Res Policy Syst* 2006.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 5. Group processes.** *Health Res Policy Syst* 2006.
- Schünemann HJ, Oxman AD, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 6. Determining which outcomes are important.** *Health Res Policy Syst* 2006.
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006.

Review

Improving the use of research evidence in guideline development: I. Guidelines for guidelines

Holger J Schünemann*¹, Atle Fretheim² and Andrew D Oxman²

Address: ¹INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ²Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Holger J Schünemann* - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no; Andrew D Oxman - oxman@online.no

* Corresponding author

Published: 21 November 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:13 doi:10.1186/1478-4505-4-13

Accepted: 21 November 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/13>

© 2006 Schünemann et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the first of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on guidelines for the development of guidelines.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We found no experimental research that compared different formats of guidelines for guidelines or studies that compared different components of guidelines for guidelines. However, there are many examples, surveys and other observational studies that compared the impact of different guideline development documents on guideline quality.

What have other organizations done to develop guidelines for guidelines from which WHO can learn?: • Establish a credible, independent committee that evaluates existing methods for developing guidelines or that updates existing ones.

• Obtain feedback and approval from various stakeholders during the development process of guidelines for guidelines.

• Develop a detailed source document (manual) that guideline developers can use as reference material.

What should be the key components of WHO guidelines for guidelines?: • Guidelines for guidelines should include information and instructions about the following components: 1) Priority setting; 2) Group composition and consultations; 3) Declaration and avoidance of conflicts of interest; 4) Group processes; 5) Identification of important outcomes; 6) Explicit definition of the questions and eligibility criteria; 7) Type of study designs for different questions; 8) Identification of evidence; 9) Synthesis and presentation of evidence; 10) Specification and integration of values; 11) Making judgments about desirable and undesirable effects; 12) Taking account of equity; 13) Grading evidence and recommendations; 14) Taking account of costs; 15) Adaptation, applicability, transferability of guidelines; 16) Structure of reports; 17) Methods of peer review; 18) Planned methods of dissemination & implementation; 19) Evaluation of the guidelines.

What have other organizations done to implement guidelines for guidelines from which WHO can learn?: • Obtain buy-in from regions and country level representatives for guidelines for guidelines before dissemination of a revised version.

- Disseminate the guidelines for guidelines widely and make them available (e.g. on the Internet).
 - Develop examples of guidelines that guideline developers can use as models when applying the guidelines for guidelines.
 - Ensure training sessions for those responsible for developing guidelines.
 - Continue to monitor the methodological literature on guideline development.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the first of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

The term guideline can be defined as "a rule or principle that provides guidance to appropriate behaviour" [1]. The Institutes of Medicine define clinical practice guidelines as "systematically developed statements to assist practitioner and patient decision about appropriate health care for specific clinical circumstances". The term "guidelines" in this document should be seen in the broad sense referring to any guideline or recommendation related to healthcare that is relevant to the mission of the WHO, including public health and health policy recommendations. A plethora of guidelines for clinical practice guidelines exist from various organizations, including national and governmental agencies and medical specialty societies. There are fewer guidelines for developing public health and health policy recommendations. We will use the term "guidelines for guidelines" as "guidelines for the development of guidelines and recommendations".

Guidelines for guidelines are important because of reports indicating that the lack of standardized guideline development leads to widely varying recommendations [2]. In this paper we addressed the following questions:

- What have other organizations done to develop guidelines for guidelines from which WHO can learn?
- What should be the key components of WHO guidelines for guidelines?
- What have other organizations done to implement guidelines for guidelines from which WHO can learn?

What WHO is doing now

An inter-cluster initiative (Guideline Development Group) led by the Evidence and Information for Policy (EIP) cluster produced the "Guidelines for WHO Guidelines" (GWG) as the recommended approach to development of WHO guidelines [3]. The process for developing the WHO document included drafting of the GWG by one group member before revision and approval by the committee. Following approval by the group, this document was reviewed and approved during a cabinet meeting before distribution as a technical cluster note to all WHO members.

The GWG (version March 10, 2003) included the following general proposals for process (see GWG "WHO documents that guide the development, dissemination and implementation of recommendations by WHO" section 5):

"b) choice of [guideline] topics; c) synthesis of the evidence; d) formulation of recommendations; e) dissemination of guidelines". The GWG makes special reference to the National Health and Medical Research Council (NHMRC) of Australia guidelines for guidelines. To accomplish proper guideline development the GWG recommends partnerships within and outside WHO according to a defined set of rules. Specific functions and composition of guideline groups are also described. The GWG also includes advice for the operationalisation of the process (section 6 GWG): a) Selection of partners; b) Organization of guideline groups; c) Process of developing guidelines; d) Guiding values. The committee also produced a self assessment checklist to ensure a consistent level of quality in the guidelines.

Although comprehensive in the coverage of topics, due to brevity most sections of the GWG could not provide the same level of detailed instructions for guideline groups to follow that other organizations provide. Moreover, it is not entirely clear to what extent WHO guideline developers adhere to the GWG, but it appears that few departments have used the GWG [4]. In part this may be a result

of a lack of a more detailed handbook that WHO guidelines development committees could follow, although there are a number of other possible explanations.

What other organisations are doing?

The use and quality of guidelines for guidelines varies across organizations that develop guidelines. There is not an accepted international standard for guideline development. However, there are several specific and detailed examples of methods adopted by other organizations. Some of these are exemplary because they give detailed guidance and resulted from a thorough process. For example, the National Institute for Health and Clinical Excellence (NICE) and the Scottish Intercollegiate Guideline Network (SIGN), both large government agencies that develop guidelines, have produced comprehensive handbooks that provide guidance for its guideline developers [5,6]. A number of professional organizations have also developed detailed guidance documents that advise their guideline developers about methods including the development of templates [7-10].

We describe the single steps of what other organizations do in regards to guideline development and on what grounds they do it in other articles in this series (see [11] for a list of articles). The steps go from setting priorities for guideline topics to implementation of the guidelines. In addition, literature has emerged from independent groups, such as the Conference on Guideline Standardization [12], that address the critical appraisal of guidelines and suggest the need for guidance for each of these steps of guideline development [13]. For example, one tool (the AGREE instrument) has demonstrated its sensitivity to differentiate higher quality guidelines that followed technical documentation from those of lower quality [14].

Methods

The methods used to prepare this review are described in the introduction to this series [11]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). As a result of prior work in the area of guideline methodology we had knowledge of existing guidelines for guidelines by organisations such as NICE, SIGN, the US Preventive Services Task Force (USPSTF), the New Zealand Guideline Group and the Australian NHRMC as well as professional societies such as the American College of Chest Physicians (ACCP) and the American Thoracic Society (ATS). We attempted to search PubMed, but were unable to devise a search strategy that was both sensitive and reasonably specific.

Given time constraints we avoided duplication with work of others and focused on sources that had systematically compiled relevant literature. We searched databases main-

tained by the Agency for Healthcare Research and Quality (AHRQ [15]), the Guidelines International Network (GIN [16]), information obtained from prominent organizations and our own files. The AHRQ database (guideline development methodology and guideline structure) is a comprehensive database that included 1205 references to both journal but also non-journal sources. The GIN database included 104 references. While there was overlap we reviewed these citations in detail and evaluated each of these references for relevance. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Findings

We found no experimental research that compared different formats of guidelines for guidelines or studies that compared different components of guidelines for guidelines. However, there are many examples, surveys and other observational studies of the impact of guideline development documents on guideline quality.

What have other organizations done to develop guidelines for guidelines that WHO can learn from?

Many large organizations that claim to develop evidence based guidelines have produced accessible, transparent and detailed guidelines for guidelines. To make a guideline for guideline credible and acceptable, individuals with expertise in methodology, process and implementation of guidelines were involved in developing a guideline for guideline document. For example, NICE involved not only various internal groups (the national collaborating centres that develop guidelines, NICE patient involvement units, etc.), but also external advisors, including individual academics and governmental institutions (e.g. SIGN) [5]. SIGN and the RAND corporation published a detailed description of the processes involved in producing guidelines for guidelines involving various stakeholders [6,17].

Other guideline developers carefully select the panels that produce guidelines for guidelines ensuring that methodologists and clinicians as well as representatives of the organization are involved. Most specialty societies have included experts and authorities in the relevant fields. While this bears the risk of involving individuals with less methodological and, therefore, relevant training for guideline development, it ensures that individuals who are knowledgeable about the relevant clinical aspects, including ongoing research, are represented and may supports buy-in by users. Aspects focusing on group processes and selection including patient representation are described in other articles in this series [18,19]. Most organizations obtain approval of the final document by a board or other governing body. While no experimental

research indicates that providing a source document (e.g. a handbook) for guideline developers improves the quality of guidelines, observational studies suggest that organizations publishing their guidelines for guidelines in the form of reference material produce more methodologically sound guidelines [20].

What should be the key components of WHO guidelines for guidelines?

We have identified 19 components that are already or should be included in the GWG and that should be described in detail in a handbook or manual for WHO guideline developers. Other reviews in this series will describe these components in greater detail. We list in parenthesis the review that describes the component in more detail and the section of the GWG that has mentioned the component. The sections in the GWG cited below often consist of a single sentence. The components are:

- 1) Priority setting ([21] and GWG 5b "Choice of topics for development of WHO guidelines")
- 2) Group composition (and consultations) ([22] and GWG 6A "Organization of guideline groups")
- 3) Declaration and avoidance of conflicts of interest ([23] and GWG 6B Note 1, annex A)
- 4) Group processes ([24] and GWG 6C3 "Process of developing guidelines")
- 5) Identification of important outcomes including cost ([25], not addressed in GWG)
- 6) Explicit definition of the question and eligibility criteria ([26,27], not addressed in GWG)
- 7) Type of study designs for different types of questions ([27], not addressed in GWG)
- 8) Identification of evidence ([27], GWG 6C2 "Undertake a systematic review")
- 9) Synthesis and presentation of evidence ([28] and GWG 5C "Synthesizing the evidence")
- 10) Specification and integration of values ([29] and GWG 6D "Guiding values")
- 11) Making judgments about desirable and undesirable effects ([29] and [30] and GWG 5d "Making recommendations")

- 12) Taking account of equity ([31], not addressed in GWG)
- 13) Grading evidence and recommendations ([30] and GWG Annex B)
- 14) Taking account of costs ([32] and GWG 5d "Making recommendations")
- 15) Applicability, transferability and adaptation of guidelines ([33] and GWG 5a "A 3 stage process")
- 16) Structure of reports ([34] and GWG 6C)
- 17) Methods of peer review ([20] and [32] not addressed in GWG)
- 18) Planned methods of dissemination & implementation ([35] and GWG 5e "Dissemination of guidelines")
- 19) Evaluation of the impact of the guideline ([36] and GWG 6C6)

What have other organizations done to implement guidelines for guidelines from which WHO can learn?

Other prominent guideline developers, such as NICE and SIGN, have ensured that those stakeholders who will become involved in guideline development also take part in the development of the guidelines for guideline. Similarly, obtaining buy-in from regions and country level representatives for GWG before agreeing on and disseminating a revised version is likely to be important. Once WHO reaches agreement on a revised version of the GWG, it should be widely disseminated and made easily available (e.g. on the Internet).

Examples and worksheets should be provided to facilitate implementation of the GWG. In addition, WHO should ensure training sessions for those responsible for developing guidelines. In their survey of 18 prominent international clinical guideline developers, Burgers and colleagues found that almost all guideline programs offer (in some organizations mandatory) training sessions to guideline developers [37]. SIGN, for example, offers a specific software program to guideline panel members and helps them with identifying specific learning needs [6]. SIGN also electronically records the amount of training of individuals who contribute to the guidelines.

The GWG should not be a static document. NICE, for example, has outlined the process for updating its guidelines for guidelines. This specifies that the formal process for updating its manual will begin three years after publication of the original manual. Interim updates may be completed to accommodate small changes outside of the

regular renewal process. NICE specifies four criteria that must all be fulfilled to qualify for a minor update: a fundamental stage in the guideline for guideline process is neither added nor removed, a fundamental methods technique or step is neither added nor removed, one or more stakeholders will not obviously be disadvantaged, and the efficiency, clarity or fairness of the process or methodology will be improved. To develop revisions of the GWG, WHO should monitor the methodological literature on guideline development and review updates of other organizations' guidelines for guidelines.

Discussion

The studies we identified and practical experience suggest that guidelines for guidelines facilitate the development of guidelines. Our review is limited in that it is not a systematic review and is based on our own judgments. However, our review has identified practical advice and components that a handbook or manual that accompany the GWG should include.

Although WHO's leadership has endorsed the GWG, a detailed handbook does not exist. Moreover, implementation of the GWG in WHO guideline programs appears to be very limited [4]. Potential reasons for this shortcoming are discussed elsewhere [4]. Key explanations include a lack of resources, technical (methodological) capacity, knowledge about the GWG, and a tradition of using non-systematic, expert opinion-based approaches [4]; as well as a lack of training and a lack of a more detailed manual.

Other organizations have invested substantial resources into guideline development, including resources to develop and implement guidelines for guidelines. Because WHO has limited resources and because well described processes used by other organizations already exist, WHO can build on existing high quality guidelines for guidelines. As we discuss in another article in this series, WHO should also consider adapting guidelines developed by other organizations, if high quality guidelines already exist [33]. In addition, WHO should consider establishing collaborations with other guideline developers to avoid unnecessary duplication of efforts and use of resources.

Further work

A systematic review of guidelines for guidelines is unlikely to yield empirical information beyond what we have found in this review, but could provide useful information about what other organisations are doing with respect to key steps in the guideline development process. We do not consider such a review to be a priority for WHO. Similar information has been obtained through surveys [20, 38, 39]. More information about specific questions regarding, for example, processes that are used

for updating, implementing and evaluating guidelines for guidelines, is more likely to come from further surveys than from a systematic review.

WHO should develop a handbook or manual that provides detailed information and examples for its guideline developers. This handbook should build on existing work, but will require time and resources. Nonetheless, this is likely necessary to improve the quality of WHO guideline development.

Efforts are needed that ensure guideline developers begin speaking the same "guideline" language and improve the standardisation of the guideline development processes used by WHO. This standardisation would help facilitate the production of guidelines that can easily be adapted to different contexts, and thus reduce global resources spent on guideline development, particularly for low and middle-income countries [33]. WHO should participate in international efforts aimed at improving guidelines for guidelines, and should aim to take a leading role in these efforts in the future.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

HJS prepared the first draft of this article. AF and ADO contributed to drafting and revising it. All authors have read and approved of the final version of this manuscript.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper. The authors thank Mr. John Fox for administrative assistance.

References

1. [www.wordreference.com: Word reference dictionary.](http://www.wordreference.com/definition) [http://www.wordreference.com/definition].
2. Beck C, Cody M, Souder E, Zhang M, Small GW: **Dementia diagnostic guidelines: methodologies, results, and implementation costs.** *J Am Geriatr Soc* 2000, **48(10)**:1195-1203.
3. **Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines. EIP/GPE/EQC/2003.1.** Geneva, World Health Organization; 2003.
4. Panisset U: **A Review of WHO recommendations published in 2005.** *Health Res Policy Syst* 2006.

5. **National Institutes for Health and Clinical Excellence** [http://www.nice.org.uk/pdf/GDM_Allchapters_0305.pdf]
6. **Scottish Intercollegiate Guideline Network. SIGN 50: A guideline developers' handbook. Section 6.** (accessed on March 17, 2006) [<http://www.sign.ac.uk/guidelines/fulltext/50/section6.html>].
7. Beyea SC: **Clinical practice guidelines--an organizational effort.** *Aorn J* 2000, **71(4)**:852, 855-8.
8. Jones RH, Ritchie JL, Fleming BB, Hammermeister KE, Leape LL: **28th Bethesda Conference. Task Force 1: Clinical practice guideline development, dissemination and computerization.** *J Am Coll Cardiol* 1997, **29(6)**:1133-1141.
9. **ATS: Attributes of ATS documents that guide clinical practice. Recommendations of the ATS Clinical Practice Committee.** *American Journal of Respiratory & Critical Care Medicine* 1997, **156(6)**:2015-2025.
10. Schünemann HJ, Munger H, Brower S, O'Donnell M, Crowther M, Cook D, Guyatt G: **Methodology for guideline development for the Seventh American College of Chest Physicians Conference on Antithrombotic and Thrombolytic Therapy: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy.** *Chest* 2004, **126**:1745-1785.
11. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006.
12. 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.
13. **Guideline development in Europe. An international comparison.** *Int J Technol Assess Health Care* 2000, **16(4)**:1039-1049.
14. 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.
15. **Agency for Healthcare Research and Quality** [<http://www.guidelines.gov>]
16. **Guidelines International Network** [<http://www.g-i-n.net>]
17. Woolf SH: **Practice guidelines: a new reality in medicine. I. Recent developments.** *Arch Intern Med* 1990, **150(9)**:1811-1818.
18. Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 3. Group composition.** *Health Res Policy Syst* 2006.
19. Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 5. Group process.** *Health Res Policy Syst* 2006.
20. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R: **Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada.** *Int J Technol Assess Health Care* 2003, **19(1)**:148-157.
21. Oxman AD, SHJ, Fretheim A.: **Improving the Use of Research Evidence in Guideline Development: Priority setting.** *Health Res Policy Syst* 2006.
22. Fretheim A, SHJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: Group composition.** *Health Res Policy Syst* 2006.
23. Boyd E, Bero L: **Improving the Use of Research Evidence in Guideline Development: 4. Managing conflicts of interest.** *Health Res Policy Syst* 2006.
24. Fretheim A, SHJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: Group process.** *Health Res Policy Syst* 2006.
25. Schünemann HJ, Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 6. Determining which outcomes are important.** *Health Res Policy Syst* 2006.
26. Oxman AD, Schünemann HJ, Fretheim A.: **Improving the Use of Research Evidence in Guideline Development: 2. Priority setting.** *Health Res Policy Syst* 2006.
27. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
28. Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006.
29. Schünemann HJ, Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 10. Integrating values and consumer involvement.** *Health Res Policy Syst* 2006.
30. Schünemann HJ, Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 9. Grading evidence and recommendations.** *Health Res Policy Syst* 2006.
31. Oxman AD, Schünemann HJ, Fretheim A.: **Improving the Use of Research Evidence in Guideline Development: 12. Incorporating considerations of equity.** *Health Res Policy Syst* 2006.
32. Edejer TTT: **Improving the Use of Research Evidence in Guideline Development: Incorporating considerations of cost-effectiveness, affordability and resource implications.** *Health Res Policy Syst* 2006.
33. Schünemann HJ, Oxman AD, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 13. Adaptation, applicability and transferability.** *Health Res Policy Syst* 2006.
34. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 14. Reporting guidelines.** *Health Res Policy Syst* 2006.
35. Fretheim A, SHJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: Disseminating and implementing guidelines.** *Health Res Policy Syst* 2006.
36. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 16. Evaluation.** *Health Res Policy Syst* 2006.
37. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15(1)**:31-45.
38. Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.** (submitted)
39. Moynihani 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.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 2. Priority setting

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim³

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway, ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ³Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 29 November 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:14 doi:10.1186/1478-4505-4-14

Accepted: 29 November 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/14>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the second of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on priority setting for health care guidelines, recommendations and technology assessments.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: There is little empirical evidence to guide the choice of criteria and processes for establishing priorities, but there are broad similarities in the criteria that are used by various organisations and practical arguments for setting priorities explicitly rather than implicitly,

What criteria should be used to establish priorities?: • WHO has limited resources and capacity to develop recommendations. It should use these resources where it has the greatest chance of improving health, equity, and efficient use of healthcare resources.

- We suggest the following criteria for establishing priorities for developing recommendations based on WHO's aims and strategic advantages:
- Problems associated with a high burden of illness in low and middle-income countries, or new and emerging diseases.
- No existing recommendations of good quality.
- The feasibility of developing recommendations that will improve health outcomes, reduce inequities or reduce unnecessary costs if they are implemented.
- Implementation is feasible, will not exhaustively use available resources, and barriers to change are not likely to be so high that they cannot be overcome.
- Additional priorities for WHO include interventions that will likely require system changes and interventions where there might be a conflict in choices between individual and societal perspectives.

What processes should be used to agree on priorities?: • The allocation of resources to the development of recommendations should be part of the routine budgeting process rather than a separate exercise.

- Criteria for establishing priorities should be applied using a systematic and transparent process.
- Because data to inform judgements are often lacking, unmeasured factors should also be considered – explicitly and transparently.
- The process should include consultation with potential end users and other stakeholders, including the public, using well-constructed questions, and possibly using Delphi-like procedures.
- Groups that include stakeholders and people with relevant types of expertise should make decisions. Group processes should ensure full participation by all members of the group.
- The process used to select topics should be documented and open to inspection.

Should WHO have a centralised or decentralised process?: • Both centralised and decentralised processes should be used. Decentralised processes can be considered as separate "tracks".

- Separate tracks should be used for considering issues for specific areas, populations, conditions or concerns. The rationales for designating special tracks should be defined clearly; i.e. why they warrant special consideration.
 - Updating of guidelines could also be considered as a separate "track", taking account of issues such as the need for corrections and the availability of new evidence.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the second of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this. In this paper we address the following questions:

- What criteria should be used to establish priorities?
- What processes should be used to agree on priorities?
- Should WHO have a centralised or decentralised process?

Questions related to group processes for committees developing guidelines and recommendations and priority setting for systematic reviews are addressed in other papers in this series [1,2].

What WHO is doing now

WHO does not have a centralised process specifically for setting priorities for the development of recommendations. A report of the Director-General to the Executive Board on WHO's strategic budgeting and planning process had this to say about priority setting in general (without specific reference to priorities for recommendations):

Specific global priorities were included in the procedural guidance for 2002–2003, and measures were provided to ensure a

shift of resources to those areas. With regard to regional and country health issues, the team received diverging views. Some staff expressed concern about the little room for specific regional or country priorities, which would not relate directly to the global priorities. Others expressed the need for flexibility during the operational planning phase, which should contribute to the achievement of global priorities in terms of reducing a health problem or improving the health status of the population. It is important to create a monitoring and evaluation system, in which such flexibility can be taken into account.

The report offered these two recommendations regarding priority setting in general:

➤ *Criteria and parameters for rationalizing the setting of programme priorities should be re-examined with the view to achieving more objectivity.*

➤ *Exercises to shift resources to priority areas should be an integral part of the programme budgeting process, and not taken up as a separate exercise.*

The Guidelines for WHO Guidelines recommends the following [3]:

Guideline development is a process which consumes resources (see Sec VII). They could be developed on almost every health topic or intervention so it is necessary for WHO to decide which topics should be given priority. It is suggested that the following areas be given priority:

- interventions that will require system changes (feasibility concerns) as opposed to those dealing solely with provider/patient interactions. WHO has greater comparative advantage in dealing with governments, for interventions which require inputs and coordination at different levels of the system. It has less comparative advantage on purely provider/patient interactions.
- cost-effective interventions that address a disease burden which is still causing major health losses, implying under-utilization of the technology (population perspective).
- interventions that are of limited or questionable effectiveness but are being used widely (opportunity costs).
- Interventions for diseases which have a high burden in developing countries, or new and emerging diseases for which there are no existing guidelines.
- interventions where there might possibly be a conflict in choices between individual and societal perspectives (political concerns: when countries will need WHO's normative support to make recommendations based on the population perspective especially in the context of other influential organizations espousing guidelines adopting an individual perspective).

The Health Evidence Network (HEN), based at the European Regional Office of WHO [4], collects policy concerns and questions from several sources and through both a proactive and a reactive approach:

Proactive:

1. Call for topics once a year, through a simple and user-friendly questionnaire to Ministries of Health of Member States, WHO technical units (TUs) including European observatory for Health care systems
2. Reviewing the work already done by HEN Members as well as their work in progress
3. Review of minutes of EU' Parliament

Reactive:

1. The Health Evidence Network To the HEN e-mailbox or direct requests from policymakers
2. Specific questions or policy concerns identified by the Observatory or WHO Technical Units in their processes of production of papers.

Once collected, all this information is translated by the HEN team into answerable questions. The list of questions is then presented to the Steering Committee once a year for its prioritization according to policy relevance, feasibility, timelines, controversy, existing evidence.

What other organisations are doing

About 30% of respondents to an international survey of producers of clinical practice guidelines or health technology assessments reported using explicit methods of setting priorities, including the use of explicit criteria, formal consensus processes, and burden of disease [4]. The US Community Preventive Services Task Force, for example,

chooses broad topics (e.g., tobacco use, cancer, diabetes, the social environment) for review on the basis of the public health burden of the problem; how preventable it is; how it relates to other public health initiatives; and the current level of research and practice activity in public health, clinical, and other settings. The agenda-setting process incorporates input from interested others.

The process of selecting specific interventions for review within those topics involves developing a candidate list of interventions, and setting priorities using a voting procedure among the team and the consultants. The Task Force approves or modifies the resulting priorities. Priority-setting criteria that are adapted for the reviews include perceived potential to reduce the burden of disease and injury; potential to increase healthy behaviors and reduce unhealthy behaviors; potential to increase the implementation of effective but not widely used interventions; potential to phase out widely used, less effective interventions in favor of more effective or more cost-effective options; and current level of interest among providers and decision makers. Other priority-setting criteria may be added as relevant and appropriate. Occasionally, review teams have engaged in formal scoring and weighting of the criteria. One or more rounds of this process results in a prioritized list of interventions.

Other respondents to the survey reported selecting topics based on consultations with their constituencies, requests from end-users, or decisions made by expert panels or a steering group without explicit methods. Seventeen of 67 respondents (25%) reported involving target users in the groups that set priorities.

Methods

The methods used to prepare this review are described in the introduction to this series [5]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register [6], the US National Guideline Clearinghouse [7],

and the Guidelines International Network [8]) for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review we searched PubMed using (clinical practice guidelines or public health guidelines) and (priority setting or setting priorities) and related articles for selected references [9,10]. We searched the Cochrane Methodology Register using priority or priorities. We reviewed the website of the 5th International Conference on Priorities in Health Care [11] and references that we had in our files [12-15].

Findings

What criteria should be used to establish priorities?

The US Institute of Medicine's (IOM) Committee on Methods for Setting Priorities for Guidelines Development in its study of setting priorities for clinical practice guidelines published in 1995 argued that the priority setting process should be open and defensible [14] They recommended six general criteria: prevalence, burden of illness, cost of managing the problem, variability in practice, potential of a guideline to improve health outcomes, and potential of a guideline to reduce costs. Because data used to make these judgements is often lacking, they suggested explicit opportunities for important unmeasured factors to be considered. They further suggested separate "tracks" for considering issues for specific populations, conditions or concerns. They argued that the rationales for designating special tracks should be defined clearly; i.e. why they warrant special consideration. They suggested that updating of guidelines should also be considered as a separate "track", taking account of issues such as the need for corrections and the availability of new evidence.

Oortwijn identified 25 criteria used to prioritise health technology assessments and categorised these into four broad categories: burden of disease, potential effects, potential costs, and uncertainty regarding application of the technology [15].

In a more recent selective review for the New Zealand Guidelines Group, the following criteria were identified as indicating that a topic is suitable for guideline development [16]:

1. *The topic is clinically important affecting large numbers of people with substantial morbidity or mortality (the burden of illness).*
2. *The topic is complex enough to initiate debate about the recommendations.*

3. *There is evidence of variation between actual and appropriate care.*

4. *There are no existing valid guidelines available to use.*

5. *There is an adequate amount of existing evidence available.*

6. *The recommendations will be acceptable to the potential users.*

7. *Implementation of the guideline is feasible, will not exhaustively use the communities' resources, and barriers to clinical change are not so high that they cannot be overcome.*

While burden of disease is commonly used as a criterion for priority setting, it should be noted that the use of summary burden of disease measures, such as disability adjusted life years (DALYs) has been criticised for focusing on disease rather than resource use and interventions, because of the assumptions about values inherent in such measures, and because of the technical limitations of such measures (see for example references [17] and [18]).

What processes should be used to agree on priorities?

Batista and Hodge in a review conducted 10 years ago found only three articles pertinent to priority setting for clinical practice guidelines [10]. They suggested the following framework for priority setting:

1. Consult with end users and other stakeholders before selecting topics.
2. Consider feasibility during the consultation.
3. Document the process used to select guideline topics.

The IOM suggested the following procedures [14]:

- the use of Delphi-like procedures for obtaining expert judgments or topic rankings through correspondence
- the use of questions that are specific, explicit and consistent with standard methods for questionnaire construction
- experimentation with more formal procedures to arrive at group judgments

They also suggested there is a need to define more narrowly and precisely topics for guideline development. They argued that this would result in more efficient organization of panels and their work, resolution of some apparent controversies, more responsive guidelines, and easier implementation.

Oortwijn identified six steps in the development of practical procedures for setting priorities [15]:

1. *Clarifying goals and responsibilities;*
2. *Choosing a general approach, method, and criteria for prioritisation;*
3. *Establishing advisory mechanisms and relations with external bodies;*
4. *Establishing arrangements to support and manage the procedure;*
5. *Defining a time table and cycle of activity; and*
6. *Evaluating and developing the procedure.*

She further identified the following ways in which approaches to priority setting can vary:

- *the extent to which the procedure is explicit and systematic*
- *the extent to which external input and advice is accepted or actively sought*
- *the relative weight given to the views of decision-makers, researchers, and others*
- *the extent to which the procedure is transparent*
- *the effort and resources devoted to the procedure*

Her main conclusion was that explicit and transparent priority setting for health technology assessment is feasible, but that some important methodological issues need to be addressed to ensure that the procedure used is valid, reliable, consistent and useful for policy making.

There is some debate, variation in practice, and limited data regarding involvement of the public in priority setting. There is limited evidence from a small survey in Australia that the public overwhelmingly want their preferences to inform priority-setting decisions [19].

Should WHO have a centralised or decentralised process?

There are two ways in which priority setting is currently decentralised: geographically (across headquarters, regional offices and countries), and across technical departments. There are limited findings in the literature to inform decisions about how this might best be handled. The IOM noted, "that it is unreasonable – indeed impossible – to expect nationally developed guidelines to cover every operational issue for every kind of setting". "Yet guidelines that leave too much to be decided at the local

level or during implementation run the risk of being ignored, misused, and modified in ways detrimental to patients." This is even more so for internationally developed guidelines. Priority setting at each level should draw on the strengths and minimize the limitations of international, national and local organizations. Thus, both centralised and decentralised processes that take account of these different strengths and limitations, as well as needs, are necessary.

Discussion

WHO has limited resources and limited technical capacity for developing recommendations. It is essential that it should set priorities for how best to use the resources and capacity it has. We did not find an empirical basis for deciding how best to set priorities. However, the use of explicit criteria and systematic processes are more likely than implicit criteria and non-systematic processes to ensure open and defensible priority setting. Based on the experience of other organisations, logic and the aims and strategic advantages of WHO we suggest that the following criteria should be used to set priorities:

- Problems associated with a high burden of illness in low and middle-income countries, or new and emerging diseases.
 - No existing guidelines or recommendations of good quality.
 - The feasibility of developing recommendations that will improve health outcomes, reduce inequities or reduce unnecessary costs if they are implemented.
 - Implementation is feasible, will not exhaustively use available resources, and barriers to change are not likely to be so high that they cannot be overcome.
 - Additional priorities for WHO include interventions that will likely require system changes and interventions where there might be a conflict in choices between individual and societal perspectives.
- The application of these criteria requires judgements. Appropriate processes are needed, in addition to explicit criteria, to ensure that these judgements are made openly, that they are taken account of in how WHO uses its resources, and that they reflect the priorities of WHO's member states, particularly those of low and middle-income countries. We suggest that the following processes be used for these reasons:
- The allocation of resources to the development of recommendations should be part of the routine budgeting process rather than a separate exercise.

- Criteria for establishing priorities should be applied using a systematic and transparent process.
- Because data to inform judgements are often lacking, unmeasured factors should also be considered – explicitly and transparently.
- The process should include consultation with potential end users and other stakeholders, including the public, using well-constructed questions, and possibly using Delphi-like procedures.
- Groups that include stakeholders and people with relevant types of expertise should make decisions. Group processes should ensure full participation by all members of the group.
- The process used to select topics should be documented and open to inspection.

Both centralised and decentralised processes should be used to take account of different strengths, limitations and needs within WHO across headquarters, regions and countries; and across different technical areas. Drawing on the suggestion of the IOM for having different tracks for considering issues for specific populations, conditions or concerns [14], we suggest:

- Both centralised and decentralised processes should be used. Decentralised processes can be considered as separate "tracks".
- Separate tracks should be used for considering issues for specific areas, populations, conditions or concerns. The rationales for designating special tracks should be defined clearly; i.e. why they warrant special consideration.
- Updating of guidelines could also be considered as a separate "track", taking account of issues such as the need for corrections and the availability of new evidence.

Further work

Many organisations are now using explicit and systematic priority setting processes for practice guidelines and health technology assessments. A more comprehensive and systematic survey of this experience could inform decisions about processes WHO should use to set priorities for recommendations. Because there is uncertainty about the best ways to set priorities, the processes that are used should be evaluated. When feasible and relevant, alternative processes should be directly compared with respect to the priorities that are generated and the resources that are used.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

1. Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 5. Group processes.** *Health Res Policy Syst* 2006.
2. Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006.
3. **Global Programme on Evidence for Health Policy.** In *Guidelines for WHO Guidelines* Geneva: World Health Organization, EIP/GPE/EQC/2003.1; 2003.
4. Moynihan R, Oxman AD, Lavis JN, Paulsen E: **Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier.** In *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.
5. Oxman AD, Fretheim A, Schünemann HJ: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
6. **Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
7. **National Guideline Clearinghouse** [<http://www.guidelines.gov>]
8. **Guidelines International Network (GIN)** [<http://www.g-i-n.net>]
9. Helou A, Perleth M, Schwartz FW: **[Determining priorities in the development of medical guidelines. 1: Criteria, procedures and actors: a methodological review of international experiences]** [Article in German]. *Z Arztl Fortbild Qualitatssich* 2000, **94**:53-60.
10. Battista RN, Hodge MJ: **Setting priorities and selecting topics for clinical practice guidelines.** *CMAJ* 1995, **153**:1233-6.
11. **5th International Conference on Priorities in Health Care** [<http://www.healthpriorities.org/>]
12. Priority-Setting Group, Council on Health Care Technology, Institute of Medicine: **National Priorities for the Assessment of Clinical Conditions and Medical Technologies. Report of a Pilot Study.** Washington DC: National Academy Press; 1990.
13. Committee on Priorities for Assessment and Reassessment of Health Care Technologies, Institute of Medicine: **Setting Priorities for Health Technology Assessment. A Model Process.** Washington DC: National Academy Press; 1992.
14. Committee on Methods for Setting Priorities for Guidelines Development, Institute of Medicine: **Setting Priorities for Clinical Practice Guidelines.** Washington DC: National Academy Press; 1995.

15. Oortwijn WJ: **First Things First. Priority Setting for Health Technology Assessment.** In (*PhD thesis*) Leiden: The Netherlands Organisation for Applied Scientific Research (TNO) Prevention and Health; 2000.
16. McKinlay E, McLeod D, Dowell T, Howden-Chapman P: **Clinical Practice Guidelines. A selective Literature Review. Prepared for the New Zealand Guidelines Group.** Wellington: Wellington School of Medicine, University of Otago; 2001.
17. Mooney G: **Discussion of the Paper: Uses of Summary Measures for Population Health by Murray, Salomon and Mathers.** Presented at WHO Conference on Summary Measures of Population Health 1999 [<http://www.who.int/whosis/smph/>].
18. Arnesen T, Kaporiri L: **Can the value choices in DALYs influence global priority setting?** *Health Policy* 2004, **70**:137-49.
19. Wiseman V, Mooney G, Berry G, Tang KC: **Involving the general public in priority setting: experiences from Australia.** *Soc Sci Med* 2004, **58**:1459.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 3. Group composition and consultation process

Atle Fretheim*¹, Holger J Schünemann² and Andrew D Oxman¹

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy

Email: Atle Fretheim* - atle.fretheim@nok.no; Holger J Schünemann - hjs@buffalo.edu; Andrew D Oxman - oxman@online.no

* Corresponding author

Published: 29 November 2006

Health Research Policy and Systems 2006, 4:15 doi:10.1186/1478-4505-4-15

Received: 07 April 2006

Accepted: 29 November 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/15>

© 2006 Fretheim et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the third of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objective: In this review we address the composition of guideline development groups and consultation processes during guideline development.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: What should be the composition of a WHO-panel that is set up to develop recommendations?

The existing empirical evidence suggests that panel composition has an impact on the content of the recommendations that are made. There is limited research evidence to guide the exact composition of a panel. Based on logical arguments and the experience of other organisations we recommend the following:

- Groups that develop guidelines or recommendations should be broadly composed and include important stakeholders such as consumers, health professionals that work within the relevant area, and managers or policy makers.
- Groups should include or have access to individuals with the necessary technical skills, including information retrieval, systematic reviewing, health economics, group facilitation, project management, writing and editing.
- Groups should include or have access to content experts.
- To work well a group needs an effective leader, capable of guiding the group in terms of the task and process, and capable of facilitating collaboration and balanced contribution from all of the group members.
- Because many group members will not be familiar with the methods and processes that are used in developing recommendations, groups should be offered training and support to help ensure understanding and facilitate active participation.

What groups should be consulted when a panel is being set up?

We did not identify methodological research that addressed this question, but based on logical arguments and the experience of other organisations we recommend that as many relevant stakeholder groups as practical should be consulted to identify suitable candidates with an appropriate mix of perspectives, technical skills and expertise, as well as to obtain a balanced representation with respect to regions and gender.

What methods should WHO use to ensure appropriate consultations?

We did not find any references that addressed issues related to this question. Based on logical arguments and the experience of other organisations we believe that consultations may be desirable at several stages in the process of developing guidelines or recommendations, including:

- Identifying and setting priorities for guidelines and recommendations
- Commenting on the scope of the guidelines or recommendations
- Commenting on the evidence that is used to inform guidelines or recommendations
- Commenting on drafts of the guidelines or recommendations
- Commenting on plans for disseminating and supporting the adaptation and implementation of the guidelines or recommendations.
- Key stakeholder organisations should be contacted directly whenever possible.
- Consultation processes should be transparent and should encourage feedback from interested parties.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the third of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Health care recommendations that are systematically and transparently developed and well informed by the best available evidence require several types of evidence and judgements. Judgements must be made about:

- The expected effects of the options that are being considered,
- Factors that might modify the expected effects in specific settings,
- Needs, risks and resources in specific settings,
- Ethical, legal and political constraints, and
- The balance between the expected benefits harms and costs if a recommendation is implemented.

A group developing recommendations must be capable of assessing the evidence that is available to inform these judgements and to make all of these different types of judgements. In this paper we address the following questions:

- What should be the composition of a WHO-panel that is set up to develop recommendations?
- What groups should be consulted when a panel is being set up?
- What methods should WHO use to ensure appropriate consultations?

Questions related to group processes or integrating values and consumer involvement are addressed in two other papers in this series [1,2].

What WHO is doing now

Expert committees are sometimes used by WHO to provide guidance. The Director General selects committee members from WHO's expert advisory panels. The members of these panels are primarily included based on "their technical ability and experience". When an expert committee is assembled it should have: "equitable geographical representation, gender balance, a balance of experts

from developed and developing countries, representation of different trends of thought, approached and practical experience in various parts of the world, and an appropriate interdisciplinary balance" [3].

Establishing an expert committee is a formal process with regulations established by the World Health Assembly. Many WHO recommendations are not developed by expert committees. Less formal procedures that are not subject to the Regulations for Expert Advisory Panels and Committees are frequently used to convene groups that develop guidelines or recommendations. "Consultations" or "proceedings" also frequently provide the basis for recommendations. Consumers or representatives of the general public are rarely included in groups that develop recommendations.

The Guidelines for WHO Guidelines state that the "Technical Guidelines Development group" should be multidisciplinary with around 8–12 individuals representing stakeholders (professionals, disease experts, primary care/public health, end users, and patients) as well as methodologists [4]. However, up to now WHO has published few recommendations that have adhered to these guidelines.

Broad consultations do not appear to be commonly used to identify potential members of expert committees or other groups that develop guidelines or recommendations. Consultations at other stages during the process of developing recommendations are also uncommon, apart from peer review of draft reports. Occasionally, draft recommendations may be circulated more widely.

What other organisations are doing

In a recent international survey of 152 units that support the use of research in developing guidelines and health policy, most respondents reported that their guidelines development panels consisted of several stakeholders or expert groups – often including end-users and consumers [5]. Another review of guidelines on hypertension and hyperlipidaemia, found that stakeholder involvement was much lower in guidelines sponsored by specialty societies than in guidelines sponsored by other groups [6]. A third review of 18 prominent guideline organisations in Australia, Canada, Europe, New Zealand and the U.S. found that guideline development groups typically consist of 10 to 20 members and the number of disciplines per group is often three to five [7]. Most of these programs invite methodological experts (epidemiologists, library scientists and others) and patient representatives. Most have permanent staff providing editorial support.

Many agencies have issued guidelines for developing guidelines [8-11] (see also our review of Guidelines for Guidelines [12]), and all such documents that we are

aware of recommend convening multi-disciplinary groups. This includes consumers, professionals working in the field, and individuals with the necessary methodological skills (e.g. epidemiologists, economists). Procedures for recruitment of panel members are usually prescribed or suggested, with emphasis on extensive consultation with relevant stakeholder-groups.

In contrast to approaches that rely heavily on clinical experts or research experts, as exemplified by many specialty societies [6] and the WHO, the National Institute for Health and Clinical Excellence (NICE) in the UK does not necessarily include experts in their guideline development groups (GDGs): "Experts attending a GDG are present because of their knowledge in a particular area. Therefore, it is important that they sit within the group and enter fully into any discussion. However, they are not full members of the group; they do not have voting rights and should not be involved in the final wording of recommendations." The role of the professional members in the GDG is to "represent the perspective(s) of the health care workers involved" [10]. In an external evaluation carried out by WHO, it was concluded that NICE is "internationally a leading agency", and that the organisation "has developed a well-deserved reputation for innovation and methodological development" [13].

Most guidelines for guidelines highlight the importance of having an effective leader. This person has a key role in facilitating "the interpersonal aspects of the group processes" and ensuring "that the group works in a spirit of collaboration, with a balanced contribution from all members" [10].

Some agencies arrange training for members of GDGs – particularly, but not solely, aimed at facilitating the active participation of consumer representatives [14]. Another suggested approach to ensure consumer involvement is to establish separate focus groups for this purpose [10].

Wide consultation in the course of developing recommendations may be done in various ways, for instance by hosting open meetings to discuss guideline drafts [8], or by posting guideline drafts on the web [10]. Peer-review is also commonly used. In the survey of 152 units that support the use of research in development of guidelines and health policy, most reported involvement of target-users in the selection of topics, e.g. in priority-setting groups, through surveys or by reviewing draft lists of priority topics [5]. Most respondents also reported having consumers involved at various stages of the development process, often by review of draft guidelines or reports [5]. In its manual for guideline developers, NICE specifies several stages during guideline development where consultations with stakeholders should take place:

- When the draft scope of the guideline has been prepared
- During the selection of panel members
- When the full draft version of the guidelines is completed

Methods

The methods used to prepare this review are described in the introduction to this series [15]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We searched PubMed and three databases of methodological literature (the Cochrane Methodology Register [16], the US National Guideline Clearinghouse [17], and the Guidelines International Network [18]) for existing systematic reviews and relevant methodological research that address these questions. We did not conduct systematic reviews ourselves. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

In our literature search on panel composition we used the terms "group composition" or "panel composition" or "consumer involvement" and "guidelines". We also checked the reference lists of key papers and contacted key researchers in the field. In our search on consultation processes, we used the term "guidelines and consultation and process".

Findings

What should be the composition of a WHO-panel that is set up to develop recommendations?

We identified relatively few articles on group composition. A key paper was a comprehensive report by Murphy and colleagues from 1998, who reviewed the research literature on group composition and clinical guideline development [19]. This systematic review identified several studies that compared recommendations by groups with different compositions, and several comparisons of judgements made by homogenous subgroups of mixed groups [19]. The authors found that "these studies, although few in number, show that differences in group composition may lead to different judgements. More specifically, members of a specialty are more likely to advocate techniques that involve their specialty." Their conclusion was that "these studies confirm that the composition of groups is important in determining the decision reached."

Knowing that groups with different compositions produce different recommendations does not necessarily tell us what group composition will provide the most appro-

appropriate recommendations. Arguments for using multidisciplinary groups are largely based on logic. For example, "Individuals' biases may be better balanced in multidisciplinary groups, and such balance may produce more valid guidelines" [20]. A report from the U.S. Institute of Medicine (IOM) put forward three arguments for multidisciplinary groups: 1) Multidisciplinary participation increases the probability that all relevant scientific evidence will be located and critically evaluated; 2) Such participation increases the likelihood that practical problems with using guidelines will be identified and addressed; 3) Participation helps build a sense of involvement or "ownership" among different audiences for the guidelines [21].

Arguments against having narrowly focused expert groups are based in part on research that have compared expert recommendations to systematic reviews and which have investigated the relationship between expertise and systematic reviews. In one study comparisons were made between recommendations of clinical experts in textbooks and major medical journals and results of meta-analyses of randomized controlled trials of treatments for myocardial infarction [22]. The investigators found that clinical experts often made recommendations that were not consistent with available research findings. Another study found strong correlations between the expertise of authors of reviews and the methods that were used in the reviews [23]. Expertise was associated with stronger opinions prior to conducting a review, less time spent conducting a review, and the use of less systematic and transparent methods.

The systematic review by Murphy and colleagues included studies of the effects that heterogeneity has on group judgement, and concluded that "The weight of evidence suggest that heterogeneity in a decision-making group can lead to a better performance than homogeneity. There is, however, also some evidence that heterogeneity may have an adverse effect because conflict may arise between diverse participants" [19].

In addition, the review found few studies of the extent to which the particular individuals that participate in a group affect the groups' decisions. The authors concluded that "the selection of individuals has some, but not a great deal, of influence on outcome", based on this limited research.

Finally, the review included research related to the optimal size of groups. The authors remark that "having more group members will increase the reliability of group judgement", while "large groups may cause coordination problems". They base their conclusion mainly on research within social and organisational psychology: "It is likely that below about six participants, reliability will decline

quite rapidly, while above about 12, improvements in reliability will be subject to diminishing returns."

We have identified one recent study not included in the review by Murphy et al. where recommendations made by groups of different composition were compared [24], as well as four studies comparing judgements made by different subgroups in mixed groups [25-28]. The findings, which are consistent with the conclusions of the review, indicate that clinical experts have a lower threshold for rating the procedures they perform as being appropriate. Another study, in which clinicians were surveyed about the appropriateness of coronary angiography for various indications, found a similar relationship [29].

We did not find any studies on the impact of group composition on public health or health systems recommendations. However, there is some evidence that suggests that the same relationships between expertise and recommendations that have been found for clinical recommendations are also found for public health recommendations; i.e. that expert recommendations are frequently not consistent with the available research evidence [30]. Moreover, those making public health and health policy recommendations may frequently not systematically consider potential adverse effects of public health and health policy interventions.

We found several systematic reviews of consumer involvement [31-35]. There are a number of relevant arguments for including consumers in groups that develop recommendations and descriptions of practical experience. However, a review of comparative studies of interventions to promote consumer involvement [31], found a "lack of evidence from comparative studies to inform decisions about desirable and adverse effects of consumer involvement in collective decisions about health care or how to achieve effective consumer involvement". The authors of another systematic review, which included non-comparative studies of how to involve consumers in setting the research and development agenda for health systems [32], recommend collaborating with "well-networked consumers and providing them with information, resources and support to empower them in key roles for consulting with their peers." They also recommended "consultations should engage consumer groups directly and repeatedly in facilitated debate." See also our review on how to integrate values and involve consumers in guideline development [2].

What groups should be consulted when a panel is being set up?

We did not identify papers addressing this question.

What methods should WHO use to ensure appropriate consultations?

We did not identify any research findings that could inform the answer to this question

Discussion

Based on the findings from the reports we have identified, there is sufficient evidence to conclude that how a panel is composed can have an important impact on conclusions drawn by a group when making recommendations for health care. In particular, clinical experts are more likely to recommend procedures linked to their own specialty than others. Furthermore, experts in a field frequently do not employ systematic methods when reviewing evidence and developing recommendations. These findings support the current recommendation in the Guidelines for WHO Guidelines: Panels should be multidisciplinary, including a broad representation of stakeholders, as well as methodologists [4].

The research evidence to guide panel composition and consultation processes is limited. However logical arguments and the experience of other organisations suggest that

- Groups that develop guidelines or recommendations should be broadly composed and include important stakeholders such as consumers, health professionals that work within the relevant area, and managers or policy makers.
- Special attention should be paid to the selection of a group leader who has a crucial role in ensuring a positive group process and that all voices within the group can be heard.
- Wide consultations should be done when selecting members of a group to develop WHO recommendations, for example by direct contact with stakeholder groups.
- Groups should include or have access to individuals with the necessary technical skills, including information retrieval, systematic reviewing, health economics, group facilitation, project management, writing and editing.
- Groups should include or have access to content experts.
- Many group members will not be familiar with the methods and processes that are used in developing recommendations, and should be offered training and support to help ensure understanding and facilitate active participation.

The process of developing recommendations, including the selection of group members, should be transparent.

The process should also include wide consultation that encourages feedback at subsequent steps in the process, which may include:

- Identifying and setting priorities for guidelines and recommendations
- Commenting on the scope of the guidelines or recommendations
- Commenting on the evidence that is used to inform guidelines or recommendations
- Commenting on drafts of the guidelines or recommendations
- Commenting on plans for disseminating and supporting the adaptation and implementation of the guidelines or recommendations.
- Commenting on what research should be conducted based on the guidelines

Further work

We have not conducted an exhaustive systematic review, but have based much of this paper on a systematic review from 1998. We have not found subsequent studies that provide conflicting evidence. There is, however, limited research for the questions addressed in this report. We do not believe that a more exhaustive review would yield a great deal of additional evidence at this time. However, it would be valuable for WHO or others to undertake and keep up-to-date systematic methodology reviews that address specific issues of group composition, including the selection of a group leader, methods for effective consultations, and methods for effective consumer involvement.

Competing interests

AF and ADO work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

AF prepared the first draft of this review. HJS and ADO contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Fretheim A, Schünemann H, Oxman AD: **Improving the use of research evidence in guideline development: 5. Group Processes.** *Health Res Policy Syst* 2006.
- Schünemann H, Fretheim A, Oxman AD: **Improving the use of research evidence in guideline development: 10. Integrating Values and Consumer Involvement.** *Health Res Policy Syst* 2006.
- Regulations for Expert Advisory Panels and Committees.** In *Resolution WHA 35 10 WHO*; 1982.
- Guidelines for WHO guidelines.** Geneva, Global Programme on Evidence for Health Policy, WHO; 2003.
- Moynihan R, Oxman AD, Lavis J, Paulsen E: **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.
- Fretheim A, Williams JW Jr., Oxman AD, Herrin J: **The relation between methods and recommendations in clinical practice guidelines for hypertension and hyperlipidemia.** *J Fam Pract* 2002, **51**:963-968.
- Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15**:31-45.
- SIGN Guideline Development Handbook.** Edinburgh, Scottish Intercollegiate Guidelines Network.; 2004.
- Handbook for the Preparation of Explicit Evidence Based Clinical Practice Guidelines.** Wellington, New Zealand Guidelines Group.; 2001.
- National Institute for Health and Clinical Excellence: **The guidelines manual.** London, National Institute for Health and Clinical Excellence; 2006.
- National Health and Medical Research Council: **A guide to the development, implementation and evaluation of clinical practice guidelines.** Canberra, National Health and Medical Research Council; 1998.
- Schünemann H, Fretheim A, Oxman AD: **Improving the use of research evidence in guideline development: I. Guidelines for Guidelines.** *Health Res Policy Syst* 2006.
- Hill S, Garattini S, van Loenhout J, O'Brien BJ, de Joncheere K: **Technology Appraisal Programme of the National Institute for Clinical Excellence.** Copenhagen, WHO Regional Office for Europe; 2003.
- Kelsom M: **The NICE Patient Involvement Unit.** *Evidence-Based Healthcare & Public Health* 2005, **9**:304-307.
- Oxman AD, Fretheim A, Schünemann H: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- National Guideline Clearinghouse** Agency for Healthcare Research and Quality 2006 [<http://www.guidelines.gov>].
- Guidelines International Network (GIN)** *Guidelines International Network* 2006 [<http://www.g-i-n.net>].
- Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T: **Consensus development methods, and their use in clinical guideline development.** *Health Technol Assess* 1998, **2**:i-88.
- Shekelle PG, Woolf SH, Eccles M, Grimshaw J: **Clinical guidelines: developing guidelines.** *BMJ* 1999, **318**:593-596.
- Clinical Practice Guidelines. Directions for a New Program.** Edited by: Field MJ and Lohr KN. Washington DC, National Academy Press; 1990.
- 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**:240-248.
- Oxman AD, Guyatt GH: **The science of reviewing research.** *Ann N Y Acad Sci* 1993, **703**:125-133.
- Raine R, Sanderson C, Hutchings A, Carter S, Larkin K, Black N: **An experimental study of determinants of group judgments in clinical guideline development.** *Lancet* 2004, **364**:429-437.
- 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.
- Vader JP, Porchet F, Larequi-Laubert T, Dubois RW, Burnand B: **Appropriateness of surgery for sciatica: reliability of guidelines from expert panels.** *Spine* 2000, **25**:1831-1836.
- 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**:513-520.
- Fitch K, Lazaro P, Aguilar MD, Martin Y, Bernstein SJ: **Physician recommendations for coronary revascularization. Variations by clinical specialty.** *Eur J Public Health* 1999, **9**:181-187.
- 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.
- Chalmers I: **Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations.** *The Annals of the American Academy of Political and Social Science* 2003, **589**:22-40.
- Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD: **Interventions for promoting consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material.** *The Cochrane Database of Systematic Reviews, Issue 3* 2006.
- Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, Gyte G, Oakley A, Stein K: **Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.** *Health Technol Assess* 2004, **8**:I-IV.
- Boote J, Telford R, Cooper C: **Consumer involvement in health research: a review and research agenda.** *Health Policy* 2002, **61**:213-236.
- Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, Tyrer P: **Systematic review of involving patients in the planning and development of health care.** *BMJ* 2002, **325**:1263.
- Coney S: **Effective Consumer Voice and Participation for New Zealand.** New Zealand Guidelines Group; 2004.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 4. Managing conflicts of interests

Elizabeth A Boyd*¹ and Lisa A Bero²

Address: ¹Department of Clinical Pharmacy, University of California, San Francisco, 3333 California Street, Suite 420, San Francisco, CA 94143-0613, USA and ²Department of Clinical Pharmacy and the Institute for Health Policy Studies, University of California, San Francisco, Suite 420, Box 0613, 3333 California Street, San Francisco, CA 94118, USA

Email: Elizabeth A Boyd* - boyde@pharmacy.ucsf.edu; Lisa A Bero - berol@pharmacy.ucsf.edu

* Corresponding author

Published: 01 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:16 doi:10.1186/1478-4505-4-16

Accepted: 01 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/16>

© 2006 Boyd and Bero; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the fourth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on conflicts of interest to answer the following questions:

1. What is the best way to obtain complete and accurate disclosures on financial ties and other competing interests?
2. How to determine when a disclosed financial tie or other competing interest constitutes a conflict of interest?
3. When a conflict of interest is identified, how should the conflict be managed?
4. How could conflict of interest policies be enforced?

Methods: We searched PubMed, the Cochrane Methodology Register and selectively searched for the published policies of several organizations. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: What is the best way to obtain complete and accurate disclosures on financial ties and other competing interests?

- Although there is little empirical evidence to guide the development of disclosure forms, minimal or open-ended formats are likely to be uninformative. We recommend the development of specific, detailed, structured forms that solicit as much information as possible about the nature and extent of the competing interests.

How to determine when a disclosed financial tie or other competing interest constitutes a conflict of interest?

- There is no empirical evidence to suggest that explicit criteria are preferable to ad hoc committee decisions when deciding if a disclosed financial tie is a conflict of interest. However, explicit criteria may make decision-making easier.

When a conflict of interest is identified, how should the conflict be managed?

- Descriptive studies suggest that appropriate management strategies are best determined on a case-by-case basis. Thus, WHO should use a wide range of management strategies to address disclosed conflicts of interest, with public disclosure of conflicts associated with each meeting as a minimum and recusal of conflicted individuals as the other extreme.

How could conflict of interest policies be enforced?

• Although there are no empirical studies of the enforcement of conflict of interest policies, descriptive studies of other organizations and institutions suggest that WHO convene a standing committee to review all financial disclosure statements prior to the commencement of committee meetings/hearings and to make management recommendations when necessary. A standard policy requiring all financial ties to be made public (i.e., recorded into the meeting minutes) should reduce the number of problematic cases. In instances where the conflicts seem intractable, a recommendation of recusal may be necessary to protect the greater interests of WHO and its constituents.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the fourth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

A conflict of interest exists when an individual's secondary interests (e.g. personal financial) interfere with or influence judgments regarding the individual's primary interests (e.g. patient welfare, education, research integrity). There is evidence demonstrating the association of financial ties with a breakdown in research integrity. Recent studies and reviews have found that industry funding for research is associated with favourable outcomes for the sponsor [1-5]. Financial ties of investigators with their sponsors (stock ownership, consulting income, etc.) are also associated with favourable research outcomes for the sponsor [5]. This scholarly evidence has been accentuated by lay media stories documenting how financial conflicts of interest have led to biased and even dangerous research (e.g., [6,7]). Biased research may be intentional or unintentional [8] and may result from damaged objectivity at multiple stages in the research process, including conceptualization of the question, design or conduct of the research, interpretation of the results, and publication (or not) of the research [9,10]. Regardless of its source, the bias associated with financial and other conflicts of interest may damage both the public's and other researcher's trust in science [11]. The type of conflict most likely to affect the public's trust is a financial conflict where the scientist tends to gain financially from a particular research outcome [11-16], although other competing interests, such as professional advancement, are important. Conflict of interest policies are designed to protect the integrity of research and decision-making processes through disclosure and transparency.

The following report relies heavily on published research related to conflicts of interest in the context of U.S. academic research and U.S. and U.K. biomedical journals because there is little empirical research from other areas.

In this paper we address the following questions:

- What is the best way to obtain complete and accurate disclosures on financial ties and other competing interests?
- How to determine when a disclosed financial tie or other competing interest constitutes a conflict of interest?
- When a conflict of interest is identified, how should the conflict be managed?
- How could conflict of interest policies be enforced?

Related questions about group composition, consultation and group processes are addressed in another paper in this series [17,18].

What WHO is doing now

Expert Advisory Panel members are currently required to disclose "all circumstances that could give rise to a potential conflict of interest as a result of their membership on an expert committee." [19]

According to the WHO Declaration of Interests for WHO Experts, a conflict of interest occurs when "the expert or his/her partner (a spouse or other person with whom s/he has a similar close personal relationship), or the administrative unit with which the expert has an employment relationships, has a financial or other interest that could unduly influence the expert's position with respect to the subject matter being considered." An apparent conflict of interest exists when the existence of an interest could result in the expert's objectivity being questioned by others, and a "potential conflict of interest exists with an interest which any reasonable person could be uncertain whether or not should be reported" [20].

The Declaration identifies 5 types of financial and other interests that must be disclosed by all experts, including proprietary interests and patents, shares or bonds in a related commercial entity, employment or consultancies, paid work or research, and grants or fellowships from a commercial entity that has an interest in the subject-matter or work of the committee. [20]

We are not aware of specific WHO documents providing guidance on how to avoid or manage conflicts of interest, and we know of no processes required to ensure that the committees discuss potential conflicts of interest on a case-by-case basis and handle them appropriately. There may be some variability in how departments collect and manage the disclosed information.

In October 2005, the WHO Office of Legal Counsel recommended a set of proposed revisions to the existing conflict of interest procedures that are similar to the recommendations in our report. These revisions would clarify the definition of a conflict of interest, include recommendations for avoiding situations that might result in conflicts of interest, and expand the relationships and affiliations that must be disclosed. The draft guidelines also recommend that a determination be made as to whether the expert's declared interest is insignificant, clearly significant, or potentially significant (para. 26). Suggestions for making this determination include weighing the nature and extent of the interest, the context of the work, and the importance of the expert's contribution (para. 29). The draft Guidelines also suggest three options for managing a conflict: 1) continue with public disclosure of the interest; 2) limit the expert's involvement; or 3) exclude the expert from the meeting or work altogether (para. 30)[21]

The draft guidelines also include a requirement that WHO experts disclose ties to the tobacco industry. This recommendation is in response to a 2000 commissioned report investigating the influence of the tobacco industry on WHO's global tobacco control policies. That report recommended that WHO formally vet prospective experts, consultants, and advisers for possible conflicts of interest related to the tobacco industry and that staff should be barred from having links with the tobacco industry [22]. In 2003, WHO's hypertension guidelines were revised in response to criticism about possible conflicts of interest among expert members [23].

What other organizations are doing

Many organizations recognize the importance of protecting against actual and potential conflicts of interest and require special employees, advisory committee members, and participants to disclose their financial ties to the organization. For example, the US Food and Drug Administration (FDA), the Cochrane Collaboration, the UK National Institute for Health and Clinical Excellence (NICE), and the US National Academies of Science all require advisory committee members and other special participants to disclose financial relationships, including research sponsorship, equity ownership, consulting fees, honoraria, related to the work or topic of the committee. These organizations use a structured disclosure form to

solicit information; they employ different standards for determining conflicts of interest and for managing them (see below).

Methods

The methods used to prepare this review are described in the introduction to this series [24] Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We searched PubMed and the Cochrane Methodology Register [25] for existing systematic reviews and relevant methodological research that address these questions. We did not conduct systematic reviews ourselves. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review, we searched PubMed for original qualitative and quantitative research using the terms "conflicts of interest" and "disclosure" and the Cochrane Methodology Register using "conflict of interest". We searched the reference lists of all relevant publications, consulted references from the Council of Medical Editors meeting on disclosure (Sept 2004) and selectively searched for the published policies of several organizations, including the Cochrane Collaboration, NICE, FDA, and National Academies of Science [26-30].

Findings

Our database searches did not yield any systematic reviews of conflict of interest or financial disclosure policies. We found several systematic reviews of literature examining the association between commercial sponsorship and outcomes favorable to the sponsor and the financial ties of investigators and favorable outcomes. We also found a number of empirical studies of particular aspects of industry involvement in science and medicine, case studies and commentaries.

What is the best way to obtain complete and accurate disclosures on financial ties and other competing interests?

We were unable to identify any randomized, controlled trials or other rigorous studies evaluating different methods for obtaining conflict of interest disclosures. Biomedical journals gather financial interest statements from authors of submitted manuscripts in three ways: 1) minimal requests about authors' professional and financial affiliations that may be perceived to have biased the presentation of results; 2) detailed instructions that request authors to describe all involvements with organizations or entities with direct financial interest in the subject matter of the study; and 3) detailed, structured checklists that require authors to declare specific interests [31]. Krimsky and others are critical of the utility of minimal and open-

ended requests [31]. Bero et al. caution that simple disclosure requests may not reveal the nature and extent to which commercial interests exert influence over the scientific process [32].

The Cochrane Collaboration Steering Group members, the Food and Drug Administration advisory committee members, NICE, and the National Academies use structured disclosure forms and request information on a range of financial ties, including research funding, paid consultancies, honoraria, equity holdings, gifts, patents, and royalties. The Cochrane Collaboration also requests information on positions of management in a related entity, including service as a director, officer, partner, trustee or employee, and information on outstanding loans from the entity. The National Academies of Science request disclosure of any position that would give the individual access to confidential information, including patient records, classified and proprietary information. NICE requests information regarding an individual's private practice that could be affected by the outcome or discussion of a particular matter or product.

There is considerable variation along other dimensions of disclosure as well. These include:

- **When** disclosures should be made: Upon appointment to the committee? Prior to the start of committee work? Under each agenda item? At the start of each committee meeting?
- **What level** of financial interest should be disclosed: Any amount (>US\$0)? Over US\$250/per year in annual income? Over US\$10,000 in equity holdings? Exact amounts or ranges (i.e., US\$1000–\$5000)?
- **What period of time** should be covered by the disclosure: The current calendar year? Past 12 months? Past 5 years? Past 5 years and future 2 years?
- **Who** should the disclosure cover: Individual only? Individual's spouse and children? Individual's institution?

How to determine when a disclosed financial tie or other competing interest constitutes a conflict of interest?

Few organizations or institutions provide explicit guidelines for determining when a particular financial relationship constitutes a conflict of interest. The Association of American Medical Colleges prohibits financial relationships between principal investigators and commercial sponsors of clinical trials, but uses a "rebuttable presumption" clause to allow the prohibition to be waived when the benefits of the research outweigh the risks of the conflict of interest [33]. The US National Institutes of Health and National Science Foundation establish financial

thresholds for disclosure – \$10,000 in annual income or 5% equity ownership in a commercial entity related to the scientific work [34].

The US FDA does not prohibit financial relationships among its Advisory Committee members and regularly issues waivers for disclosed conflicts of interest when 1) "the disqualifying financial interest is not so substantial that it is likely to affect the integrity of an employee's services to the government;" and 2) the "need for the employee's services outweighs the potential conflicts of interest" [29]. In making these determinations, the FDA evaluates "the type of interest creating the disqualification; the identity of the person whose financial interest is at issue; the dollar value of the disqualifying financial interest including its value in relationship to the individual's overall assets; the nature and importance of the individual's role in the matter, including the extent to which the employee is called upon to exercise discretion; the sensitivity of the matter; and the need for the employee's services in the particular matter" [29].

These criteria are in line with the criteria used by University of California conflict of interest committees. In the only empirical study to date of how conflict of interest committees define and manage disclosed financial relationships of faculty investigators [35], found that committees typically examined the nature of the proposed scientific work (basic or applied), the overlap between paid activities and the research topic, the length and dollar amount of the relationship between the investigator and the commercial entity, and the degree to which the investigator could be seen as independent of the company's interests.

The overall lack of explicit criteria for determining which relationships constitute conflicts of interest reflect a common perception that these decisions should be made on an ad hoc basis and that the organization must always balance its own needs for the particular expertise of the individual with the needs of the public (in terms of advancing scientific discovery as well as trust in the scientific process). Little is known, however, about the needs and understandings of the public in this regard. The few studies we have identified to date provide evidence of both favorable and unfavorable reactions of the public [36,37]. Professionals with industry ties are more supportive of financial ties than those without industry ties. Investigators recognize general risks of conflicts of interest, but not for themselves. Investigators tend to support disclosure of financial ties, although there is evidence that disclosure leads to more critical review of research findings [38]. Schroter and colleagues showed that the overall importance, relevance, validity, and believability of studies disclosing

competing interests were rated lower by readers than those without competing interests. [39]

When a conflict of interest is identified, how should the conflict be managed?

The only empirical studies of management decisions in conflicts of interest detail a number of management strategies that are commonly used by university conflict of interest committees [35,40]. These possible management strategies include: disclosure of the financial tie(s) in publications and public presentations; reducing equity holdings to below 5%; altering consulting agreements to ensure separation between consulting and research work; eliminating the financial tie; appointing oversight committees to review the scientific process and resulting research; and recusal. Government and professional society guidelines recommend that institutions "manage" the financial conflicts of interest of their researchers. Disclosure of financial ties in all publications and presentations is the most frequently used management strategy [40-43].

Scientific journals are also encouraging disclosure as a way of dealing with financial conflicts of interest [44], however, the adequacy of disclosures in scientific articles has been questioned [32,45]. Even when financial sponsorship is disclosed, few studies describe the role of the sponsor [46]. A study of the relationships between authors of clinical practice guidelines and the pharmaceutical industry found considerable interaction between guideline authors and the pharmaceutical industry [47]; another study found that clinical practice guidelines published in journals almost never published conflict of interest statements along with the guidelines [48].

The FDA in 2002 issued draft guidance amending their disclosure regulations related to Advisory Committee members. The draft guidance now requires that Advisory Committee members granted waivers of their conflicts of interest will have the nature and magnitude of their conflicts of interest disclosed and read into the public record at the start of the committee hearings. [28]. NICE, which is currently reviewing its policies on disclosure and conflicts of interests, recommends that "members should declare all interests at the beginning of all appraisals" and that those declarations of interests be kept in files available for public scrutiny or are recorded in the minutes of the meeting [27]. The Cochrane Collaboration publishes the declarations of interests of its Steering Group members [26].

Although disclosure of financial ties is becoming more accepted within the scientific and policy communities, there are widely varying opinions about the adequacy of disclosure as a management strategy for financial conflicts of interest. Some critics of disclosure feel that it is unne-

cessary and can taint the reputation of "good" researchers [49,50]. Others believe that "the key to avoiding conflict of interest is public disclosure" [51]. Studies that disclose industry sponsorship have a systematic bias towards outcomes that favor the sponsor [3,5,52,53], so, therefore, disclosure does not eliminate bias. Although disclosure does not eliminate the association of research funding with outcomes favorable to the sponsor, many argue that it can minimize perceived conflicts of interest.

Additional research is necessary to be able to evaluate different methods for defining conflicts of interest and to determine their relative impact on the decision-making capabilities of the organization.

How should conflict of interest policies be enforced?

There is no empirical evidence evaluating the enforcement of conflict of interest policies. Most organizations and academic institutions convene a standing or ad hoc committee to review financial interest disclosures and, where deemed necessary, recommend management strategies. The US FDA vets all financial disclosure statements through a multi-stage process, beginning with initial review, followed by consultation with the individual and an FDA official, review by the FDA Ethics staff, and final approval by the appointing official. The FDA operates under federal regulations and thus has the power to enforce its decisions [28,29]. The Cochrane Collaboration directs unclear cases of financial disclosure for reviews to a "Funding Arbiter" who convenes a standing panel of four to give guidance [26].

Further work

There is currently a lack of empirical evidence regarding the most effective ways to determine the existence of conflicts of interest, manage conflicts of interest, and enforce conflict of interest policies. Additional research is necessary to evaluate different methods for defining conflicts of interest and to determine their relative impact on the decision-making capabilities of the organization. WHO's proposed draft recommendations (October 7, 2005) represent a more rigorous evaluation of conflict of interest because it requires more complete disclosure, clearer standards for evaluating conflicts of interest, and explicit management strategies.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

EAB prepared the first draft and revisions of this paper. LAB contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge the members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper.

References

- Barnes DE, Bero LA: **Why review articles on the health effects of passive smoking reach different conclusions.** *Jama* 1998, **279(19)**:1566-1570.
- Barnes DE, Bero LA: **Scientific quality of original research articles on environmental tobacco smoke.** *Tob Control* 1997, **6(1)**:19-26.
- Bekelman JE, Li Y, Gross CP: **Scope and impact of financial conflicts of interest in biomedical research: A systematic review.** *JAMA* 2003, **289(454-65)**.
- Cho MK, Bero LA: **The quality of drug studies published in symposium proceedings.** *Ann Intern Med* 1996, **124(5)**:485-489.
- Lexchin J, Bero LA, Djulbegovic B, Clark O: **Pharmaceutical industry sponsorship and research outcome and quality: systematic review.** *Bmj* 2003, **326(7400)**:1167-1170.
- Wilson D, Heath D: **He saw the tests as a violation of 'trusting desperate human beings'.** In *The Seattle Times* Seattle ; 2001.
- Wilson D, Heath D: **Uninformed consent: they call the place Mother Hutch.** In *The Seattle Times* Seattle ; 2001.
- Dana J: **Harm avoidance and financial conflict of interest.** *Journal of Medical Ethics* 2003, **Online Electronic Version**:1-18.
- Bero LA: **Tobacco industry manipulation of research.** *Public Health Rep* in press.
- Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ, Guyatt GH: **Users' guide to detecting misleading claims in clinical research reports.** *BMJ* 2004, **329**:1093-1096.
- Friedman PJ: **The impact of conflict of interest on trust in science.** *Science and Engineering Ethics* 2002, **8(3)**:413-420.
- Drazen JM, Koski G: **To protect those who serve [editorial].** *The New England Journal of Medicine* 2000, **343**:1643-1645.
- Baltimore D: **The worsening climate for biological research; is the public losing faith in science?** *Technology Review* 1989, **92(4)**:22.
- Haerlin B, D. P.: **How to restore public trust in science [commentary].** *Nature* 1999, **400**:499.
- Cohen JJ: **Trust us to make a difference: Ensuring public confidence in the integrity of clinical research.** *Academic Medicine* 2001, **76(2)**:209-214.
- DeAngelis CD: **Conflict of interest and the public interest.** *Journal of the American Medical Association* 2000, **284(17)**:2237-2238.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 3. Group composition and consultation process.** *Health Research Policy and Systems* 2006.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 5. Group processes.** *Health Research Policy and Systems* 2006.
- World Health Organisation: **Regulations for Expert Advisory Panels and Committees.** 2003:101-109.
- World Health Organisation: **Global Programme on Evidence for Health Policy, Guidelines for WHO Guidelines.** 2003:15.
- World Health Organisation: **Draft Guidelines for Declaration of Interest (WHO Experts).** 2005.
- Godlee F: **Tobacco industry's smoke screen blown.** *The Scientist* 2000.
- World Health Organisation: **WHO/ISH Hypertension Guidelines.** 2003.
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006.
- Cochrane Methodology Register [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- Cochrane Collaboration: **Cochrane Collaboration policy on commercial sponsorship.** 2004.
- National Institute for Clinical Excellence: **Review of code of practice on declaration of conflicts of interest.** .
- Draft guidance on disclosure of conflicts of interest for special government employees participating in FDA product special advisory committees.** U.S. Food and Drug Administration; 2002.
- Policies and procedures for handling conflicts of interest with FDA advisory committee members, consultants, and experts.** U.S. Food and Drug Administration; 2005.
- Policy on committee composition and balance and conflicts of interest for committees used in the development of reports.** The National Academies; 2003.
- Krimsky S: **Science in the Private Interest.** Lanham, Maryland , Rowan & Littlefield; 2003:247.
- Bero LA, Glantz S, Hong MK: **The limits of competing interest disclosures.** *Tobacco Control* 2005, **14**:118-126.
- Association of American Medical Colleges: **Protecting subjects, preserving trust, promoting progress: Guidelines for developing and implementing a policy concerning individual financial interests in human subjects research.** Philadelphia ; 2001.
- U.S. Department of Health and Human Services: **Objectivity in research.** Volume 35810 (codified at 42 CFR 50. 60 Federal Register; 1995.
- Boyd E, Lipton S, Bero L: **Implementation of financial disclosure policies to manage conflicts of interest.** *Health Affairs* 2004, **23(2)**:206-214.
- Sprecker K: **How involvement, citation style, and funding source affect the credibility of university scientists.** *Science Communication* 2002, **24(1)**:72-97.
- Kim SYH, et al.: **Potential research participants' views regarding researcher and institutional financial conflicts of interest.** *Journal of Medical Ethics* 2003. Online Electronic Version: p. 1-18
- Glaser B, Bero L: **Attitudes of Academic and Clinical Researchers toward Financial Ties in Research.** *Science and Engineering Ethics.* 2005, **11**:553-573.
- Schroter S, Morris J, Chaudhry S, Smith RD, Barratt H: **Does the type of competing interest statement affect readers' perceptions of the credibility of research? A randomized trial.** *BMJ* 2004, **328**:742-743.
- Boyd EA Ph.D., Bero LA Ph.D.: **Assessing Faculty Financial Relationships with Industry.** *Journal of the American Medical Association* 2000, **284(November 1, 2000)**:2209-2238.
- Boyd EA, Cho MK, Bero LA: **Financial conflict-of-interest policies in clinical research: issues for clinical investigators.** *Academic Medicine* 2003, **78(8)**:769-774.
- Krimsky S, Rothenberg, LS., Stott, P., Kyle, G.: **Financial interests on authors in scientific journals: a principal investigator lot study of 14 publications.** *Science and Engineering Ethics* 1996, **2(4)**:395-410.
- Shamoo AE, Resnik DB: **Responsible Conduct of Research.** New York , Oxford University Press; 2003:345.
- Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication.** International Committee of Medical Journal Editors; 2005. Available at: <http://www.icmje.org>
- Krimsky S, Rothenberg LS: **Conflict of interest policies in science and medical journals: Editorial practices and author disclosures.** 2001, **7**:205-218.
- Gross CP, Gupta AR, Krumholz HM: **Disclosure of financial competing interests in randomised controlled trials: cross sectional review.** *BMJ* 2003, **326**:526-527.
- Choudhry NK, Stelfox HT, Detsky AS: **Relationships between authors of clinical practice guidelines and the pharmaceutical industry.** *JAMA* 2002, **287(5)**:612-617.
- Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DS, Haidich AB, Giannakakis IA, Ioannidis JPA: **Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions.** *BMC Medical Research Methodology* 2001, **1(3)**.
- Jansen LA, Sulmasy DP: **Bioethics, conflicts of interest, and the limits of transparency.** *Hastings Center Report* 2003, **33**:40-43.
- Rothman K: **Conflict of interest: The new McCarthyism in science.** *JAMA* 1993, **269**:2782-2784.
- Duderstadt J: **A university for the 21st century.** Ann Arbor , University of Michigan Press; 2000.
- Barnes D, Bero L: **Why review articles on the health effects of passive smoking reach different conclusions.** *Journal of the American Medical Association* 1998, **279**:1566-1570.
- Barnes D, Bero L: **Scientific quality of original research articles on environmental tobacco smoke.** *Tobacco Control* 1997, **6**:19-26.

Review

Improving the use of research evidence in guideline development: 5. Group processes

Atle Fretheim*¹, Holger J Schünemann² and Andrew D Oxman³

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway, ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ³Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Atle Fretheim* - atle.fretheim@nokc.no; Holger J Schünemann - hjs@buffalo.edu; Andrew D Oxman - oxman@online.no

* Corresponding author

Published: 01 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:17 doi:10.1186/1478-4505-4-17

Accepted: 01 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/17>

© 2006 Fretheim et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the fifth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objective: In this review we address approaches to facilitate sound processes within groups that develop recommendations for health care.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key question and answer: What should WHO do to ensure appropriate group processes?

Various strategies can be adopted to ensure that the group processes in play when panels are developing recommendations are inclusive, so that all voices can be heard and all arguments given fair weight, including

- the use of formal consensus development methods, such as the Nominal Group Technique or the Delphi method
- the selection of a group leader who is qualified and responsible for facilitating an appropriate group process.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the fifth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

A group that is convened to formulate recommendations will necessarily enter some sort of consensus development process. The panel-members will pass judgements on the available research evidence, consider various trade-offs (between expected benefits, harms and costs), and finally try to reach a consensus on what recommendation to make. The consensus may be reached through informal processes or more formal methods.

In this paper we address the following question: What should WHO do to ensure appropriate group processes? Questions related to group composition, integrating values and consumer involvement are addressed in other papers in this series [1,2].

What is WHO doing now?

We are not aware of any examples of the use of formal consensus development methods by groups that have developed recommendations on behalf of WHO. In documents describing procedures for Expert Committees, it is stated that the meetings "shall normally be of private character" [3].

In our literature search we identified a paper describing the use of formal consensus development methods in modifying WHO's "Guidelines for the management of HIV/AIDS in adults and children" for use in Malawi and Barbados [4]. The method employed was the Nominal Group Technique.

What are other organisations doing?

Several formal approaches for reaching consensus exist, and some organisations use these in the development of

clinical practice guidelines. Three of the most common methods for reaching consensus are the Nominal Group Technique (NGT), the Delphi Method, and Consensus Conferences. A brief description of commonly used consensus development methods is found in Table 1. However, there is considerable variation in how these techniques are implemented in practice.

In a recent international survey of organisations that develop clinical practice guidelines or health technology assessments, approximately 42 % of the respondents reported using formal consensus development methods [5]. A smaller survey of prominent guideline developers reported that 7 of 18 programs used formal consensus methods to formulate recommendations [6].

With informal consensus development, a strategy is needed to ensure appropriate group processes. Typically responsibility for this is given to the group leader. Consequently, much weight is put on selecting the right person for this position. The National Institute for Health and Clinical Excellence (NICE) in the UK notes that the group leader "needs to allow sufficient time for all members to express their views without feeling intimidated or threatened and should check that all the members in the groups agree to endorse any recommendations" [7]. Furthermore, "The Chair should be selected as someone who is neutral and who has enough expertise in coordinating groups of health professionals and patients/carers so that the appointment is acceptable to all." [7].

Methods

The methods used to prepare this review are described in the introduction to this series [8]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We searched PubMed and three databases of methodological literature (the Cochrane Methodology Register [9], the US National Guideline Clearinghouse [10] and the Guidelines International Network [11]) for existing systematic reviews and relevant methodological research that address these questions. We did not conduct systematic reviews ourselves.

Table 1: Characteristics of various consensus development methods (from Murphy et al. [12])

Consensus development method	Mailed questionnaires	Private decisions elicited	Formal feedback of group choices	Face-to-face contact	Interaction structured	Aggregation method
Informal	No	No	No	Yes	No	Implicit
Delphi method	Yes	Yes	Yes	No	Yes	Explicit
NGT	No	Yes	Yes	Yes	Yes	Explicit
RAND version	Yes	Yes	Yes	Yes	Yes	Explicit
Consensus development conference	No	No	No	Yes	No	Implicit
Other methods						
Statisised group	No	Yes	No	No	-	Explicit
Social judgement analysis	No	Yes	Yes	Yes	No	Implicit
Structured discussion	No	No	No	Yes	Yes	Implicit

The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

In the literature search we used the term "consensus and process and method". We also checked the reference lists of key papers and contacted researchers in the field.

Findings

What should WHO do to ensure appropriate group processes?

One comprehensive review on consensus development methods was identified, and it provided most of the key findings for this report [12], together with an updated review that largely confirms the findings [13]. The review addresses three questions related to interaction within guideline development groups: 1) Does the choice of consensus development method influence the group's decision? 2) Does the setting for the group meetings affect the consensus decision? 3) Do the characteristics of a group facilitator affect the consensus decision?

Various measurements of decision quality were used as outcome measures in the comparative studies that were included in the review. For instance, comparison with "gold standard", such as asking the groups to reach an agreement on "questions that have correct answers which the participants do not know with any precision," e.g. "What is the diameter of Jupiter?". For ranking tasks (e.g. "to rank items in terms of their value for survival on the moon"), the group decision "can be compared with rankings by experts." The applicability of these types of studies for processes taking place within guideline development groups is not obvious.

For choice of consensus development methods, the reviewers identified 16 studies comparing NGT with informal methods, 11 comparing the Delphi method with informal methods, and seven studies comparing NGT and Delphi. Interpreting the results is not straight-forward since "the studies also differ in the particular way they operationalise the method used". The reviewers did not find any comparative studies involving consensus development conferences. Their summary conclusion was that "Formal methods generally perform as well or better than informal methods, but it is difficult to tell which of the formal methods is best."

With regards to the settings for group meetings, the reviewers concluded that "There is little research which actually looks at this question. However, of the many factors which can influence decision-making, except for extreme environments, the environment is likely to have only a marginal impact."

Concerning characteristics of a group facilitator, the research base is difficult to interpret as "the models of leadership used are often not directly transferable to facilitation". Although there is "very little work that looks at the effects of facilitation on group decision-making", the reviewers believe that "it is likely that this key role will influence group decision-making."

We identified one additional study that compared informal consensus with a formal consensus method ("the appropriateness method") for developing clinical practice guidelines on the management of low-back pain [14]. The investigators found that guideline statements resulting from the two approaches were "qualitatively similar", however the formal method produced statements that in some instances were "more clinically specific".

Discussion

The idea of bringing people together to develop recommendations is based on the understanding that they all have something to contribute. Thus, it is essential to secure that all participants can be heard and have the opportunity of influencing the outcome of the process. This is a common understanding among groups that develop guidelines, and many have therefore adopted specific strategies to ensure appropriate group processes.

Given the costs of group meetings, different languages and cultural differences, it is especially important for WHO to ensure that all of the invited members contribute fully to the development of recommendations. Transparency is important to ensure that groups know and adhere to the methods that they are supposed to be use. For instance, the group may report that they base their recommendations on research evidence, while they in reality reach their conclusions on a different basis. A qualitative study of decision-making processes within drug-selection committees in hospitals in the UK, for example, found that many decisions were not based on research findings, despite being reported as if they were: "reports of decisions...are written so as to account for the decision in terms of scientific rationality...rather than the local rationality that was actually employed" [15].

The research base to inform the choice of strategy to ensure appropriate group processes is limited, however in addition to logical arguments there is also some empirical evidence in support of using formal consensus development methods rather than relying only on informal processes. Having a group leader that facilitates the group process is likely essential. Conflicts may arise within groups and the leader of the group will have an important role in trying to manage these. Dealing with conflict is usually a difficult task, and WHO should consider establishing routines to support groups in managing these.

A weakness of our review was the literature search was limited to PubMed and three databases of methodological literature and did not include additional searches in the social science literature.

Further work

In general, there is need for research to learn more about the relative merits of various methods for facilitating sound group processes. Head to head comparisons of different consensus development methods within groups that develop recommendations for health care should be done, since most research so far has taken place in very different settings. Also, research is needed to identify the most critical selection criteria and processes for selecting a chairperson for groups developing recommendations.

Competing interests

AF and ADO work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapies Guidelines.

Authors' contributions

AF prepared the first draft of this review. HJS and ADO contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Fretheim A, Schünemann H, AD O: **Improving the use of research evidence in guideline development: 3. Group Composition and Consultation Process.** *Health Res Policy Syst* 2006.
- Schünemann H, Fretheim A, AD O: **Improving the use of research evidence in guideline development: 10. Integrating Values and Consumer Involvement.** *Health Res Policy Syst* 2006.
- Regulations for Expert Advisory Panels and Committees.** In *Resolution WHA 35.10 WHO*; 1982.
- Wabitsch R, Margolis CZ, Malkin JE, Abu-Saad K, Waksman J: **Evaluating the process of tailoring the global HIV/AIDS practice guidelines for use in developing countries.** *Int J Qual Health Care* 1998, **10**:147-154.
- Moynihan R, Oxman AD, Lavis J, Paulsen E: **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.
- Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15**:31-45.
- National Institute for Health and Clinical Excellence: **The guidelines manual.** London, National Institute for Health and Clinical Excellence; 2006.
- Oxman AD, Fretheim A, Schünemann S: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- National Guideline Clearinghouse** Agency for Healthcare Research and Quality 2006 [<http://www.guidelines.gov>].
- Guidelines International Network (GIN)** *Guidelines International Network* 2006 [<http://www.g-i-n.net>].
- Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T: **Consensus development methods, and their use in clinical guideline development.** *Health Technol Assess* 1998, **2**:i-88.
- Hutchings A, Raine R: **A systematic review of factors affecting the judgments produced by formal consensus development methods in health care (In press).** *J Health Serv Res Policy* 2006.
- Shekelle PG, Schriger DL: **Evaluating the use of the appropriateness method in the Agency for Health Care Policy and Research Clinical Practice Guideline Development process.** *Health Serv Res* 1996, **31**:453-468.
- Jenkins KN, Barber N: **What constitutes evidence in hospital new drug decision making?** *Soc Sci Med* 2004, **58**:1757-1766.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 6. Determining which outcomes are important

Holger J Schünemann*¹, Andrew D Oxman² and Atle Fretheim²

Address: ¹INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ²Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Holger J Schünemann* - hjs@buffalo.edu; Andrew D Oxman - oxman@online.no; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 01 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, **4**:18 doi:10.1186/1478-4505-4-18

Accepted: 01 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/18>

© 2006 Schünemann et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the sixth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on determining which outcomes are important for the development of guidelines.

Methods: We searched five databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct a complete systematic review ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We did not find a systematic review that addresses any of the following key questions and we found limited relevant research evidence.

What methods should WHO use to identify important outcomes?

- Methods of outcome identification should be transparent and explicit.
- The consultation process should start with identification of all relevant outcomes associated with an intervention.
- Those affected, including consumers, should be involved in the selection of outcomes.
- A question driven approach (what is important?) is preferable to a data driven approach (what data are at hand?) to identify important outcomes.

What type of outcomes should WHO consider and how should cultural diversity be taken account of in the selection of outcomes?

- Desirable (benefits, less burden and savings) and undesirable effects should be considered in all guidelines.
- Undesirable effects include harms (including the possibility of unanticipated adverse effects), greater burden (e.g. having to go to the doctor) and costs (including opportunity costs).
- Important outcomes (e.g. mortality, morbidity, quality of life) should be preferred over surrogate, indirect outcomes (e.g. cholesterol levels, lung function) that may or may not correlate with patient important outcomes.
- Ethical considerations should be part of the evaluation of important outcomes (e.g. impacts on autonomy).

- If the importance of outcomes is likely to vary across cultures, stakeholders from diverse cultures should be consulted and involved in the selection of outcomes.

How should the importance of outcomes be ranked?

- Outcomes should be ranked by relative importance, separated into benefits and downsides.
 - Information from research on values and preferences should inform the ranking of outcomes whenever possible.
 - If the importance of outcomes is likely to vary across cultures, ranking of outcomes should be done in specific settings.
 - If evidence is lacking for an important outcome, this should be acknowledged, rather than ignoring the outcome.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the sixth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

An outcome can be defined as a measure of an intervention's desirable (benefits, less burden and savings) or undesirable effects (including harms, greater burdens and cost). Those making health care recommendations always should consider the benefits, potential harms, including the potential for unanticipated adverse effects, burdens (e.g. having to take a pill), and costs, including opportunity costs. Identifying all known and plausible outcomes that are important to those affected and associated with an intervention is a key step in formulating questions for guideline development. Unfortunately, guideline developers sometimes select outcomes based on what has been assessed in studies rather than based on what is important to those affected.

Since interventions affect several outcomes (e.g. some hypertensive treatments have effects on mortality, stroke, diabetes, libido), guideline developers need to consider their relative importance. This is also true for public health and health systems interventions. For example, media campaigns might cause anxiety as well as promoting a desired health behaviour, and there are always associated costs. At the very least, there are opportunity costs.

Patients may assign different values to outcomes than clinicians and clinical experts involved in guideline development [1]. In addition, surrogate outcomes such as laboratory measures that are part of the clinician's repertoire often do not correlate with patient important outcomes and guideline developers should scrutinize surrogate outcomes about how directly they relate to patient important outcomes.

In this paper we addressed the following questions:

- What methods should WHO use to identify important outcomes?
- What type of outcomes should WHO consider and how should cultural diversity be taken account of in the selection of outcomes?
- How should the importance of outcomes be ranked?

Questions related to integrating values and consumer involvement are specifically addressed in another paper in this series [2].

What WHO is doing now

The Guidelines for WHO Guidelines suggests the following:

- "To identify the issues to be addressed, it is helpful to develop a logic and analytical frameworks guide (Woolf, 1994)" [3]. (GWG 6C1 Process of developing guidelines)
- "Spell out any tradeoffs between the cost of applying possible recommendations on a population basis, and the population health impacts" in the second stage of guideline development. (section 5d Making recommendations)
- "All evidence, including that on safety, should be clearly laid out in an evidence table" (GWG section 6C2).

Despite these guidelines, a review of several WHO guidelines (e.g., contraceptive use, hypertension, air pollution, inpatient treatment of malnourished children, treatment of non-breastfed children) revealed that the process of outcome identification is usually not described.

What other organisations are doing

The UK National Institute for Health and Clinical Excellence (NICE) defines a very explicit process for the identification of outcomes using the Population, Intervention, Comparison and Outcome (PICO) format for the devel-

opment of questions [4]. The NICE handbook asks guideline panels to consider:

- What outcome is really important for the patient?
- Which outcomes should be considered: intermediate or short-term measures (e.g., mortality, morbidity and treatment complications, quality of life, cost, etc)?

Similarly, the National Health and Medical Research Council of Australia bases its approach on the NICE handbook and defines the appropriateness of the outcomes by asking "Are they relevant to the patient?" [5,6].

SIGN underlines (section 5.1. of the SIGN handbook for guideline developers) that patients' perspectives should be included early in the guideline development process [7]. Therefore, SIGN prescribes to conduct a specific literature search designed to cover both quantitative and qualitative evidence about outcomes without limitations of study design, but this is not (yet) done consistently (Robin Habour, personal communication). In theory, the results of this search inform the development of key questions. SIGN uses the PICO format for question development.

The United States Preventive Services Taskforce (USPSTF) [8] describes that value judgments are involved in using the information in an outcomes table to rate either benefits or harms. USPSTF uses a 4-point scale to rate importance. Value judgments are also needed to weigh benefits against harms and to arrive at a rating of net benefit. The USPSTF does not use formal processes for identifying outcomes as part of the question formulation. Specialty societies do not consistently acknowledge a formal process for question development and the processes are often not transparent.

Methods

The methods used to prepare this review are described in the introduction to this series [9]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We reviewed existing guidelines for guidelines to identify processes for outcome identification and ranking. We also searched PubMed using (guideline OR policy making) and (identification) and (outcomes) as search terms (MESH headings/keywords) for systematic reviews and studies of methods for identifying outcomes for guideline development (69 citations). We also searched the Cochrane Methodology Register and Database of Methodology Reviews using the keywords "outcome" and "identification". We also searched databases maintained by the Agency for Healthcare Research and

Quality (AHRQ, [10]) and the Guidelines International Network (GIN, [11]). These searches were supplemented with information obtained directly from guideline development organizations and our own files. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organizations are doing and logical arguments.

Findings

We did not find a systematic review that addresses any of the key questions and we found very little relevant research evidence.

What methods should WHO use to identify important outcomes?

Few guideline developers have included descriptions of methods for the identification of important outcomes. SIGN uses an approach that begins with conducting a search for evidence using the patient perspective before finalizing the formation of the question. Most other guideline developers have not described formal processes of identifying important outcomes when formulating guideline questions. To be reproducible and understandable, the methods of outcome identification should be transparent and explicit.

NICE suggests facilitating the process of formulating questions, "it may be helpful to construct a diagram listing outcomes and other key criteria the [guideline] group has considered important. Once the question has been framed, key words can be identified as potential search terms" [12]. NICE involves patient organizations in developing guideline scopes and routinely includes at least two patient or caregiver members who provide a patient perspective on all guideline development activities including the formulation of clinical questions and defining of relevant outcomes [13]. Owens and Nease suggest the use of influence diagrams to identify important outcomes and focus guideline questions [14]. They argue this helps to delineate an explicit link between interventions and outcomes, shifts the focus from broad questions to more sharply delineated questions to be addressed, and highlight the importance of a clear, unambiguous statement of whose benefit, downsides and costs are under consideration. Thus, this limited evidence suggests that a consultation process should start with identification of all relevant outcomes associated with an intervention.

Bravata and colleagues. conducted an overview of reviews to identify innovative methods for question formulation related to challenging topics in health care (organization, delivery and financing of health care) [15]. They found that the use of decision analytical frameworks for defining a question and systematic methods such as influence diagrams influenced how questions were formulated. Fur-

thermore, systematic methods appeared to have an effect on search strategies to identify underlying evidence.

There is little empirical evidence to inform decisions about what methods to use to identify important outcomes. Given the paucity of data on patients' and the public's values WHO should consider using other evidence such as systematic summaries and original research on people's perspectives and experiences ("views" studies) alongside trials of effectiveness. Summarizing views studies in a systematic way could lead to a greater breadth of perspectives and a deeper understanding of public health issues from the point of view of those targeted by interventions. Harden et al. suggest that this methodology is likely to create greater opportunities for people's own perspectives and experiences to inform policies to promote their health [16].

Compared to the limited evidence about methods to identify important outcomes in guidelines, there is a large literature that documents that the importance of outcomes can vary within and across cultures, and between health care professionals and patients [1,17]. In addition, both clinical and public health interventions can have effects that are important to consumers, but are not considered important by researchers or health professionals in part because values differ between decision makers. This suggests two key elements of any approach that is used to identify important outcomes. First, all relevant stakeholders (including consumers) should be consulted at an early stage in the process. Secondly, the formulation of questions and the search for evidence should then consider all relevant outcomes.

What type of outcomes should WHO consider and how should cultural diversity be taken account of in the selection of outcomes?

The AGREE Collaboration states that the guidelines development process "involves taking into account the benefits, harms and costs of the recommendations, as well as the practical issues attached to them" [18]. The AGREE instrument suggests guidelines "should consider health benefits, side effects, and risks of the recommendations. For example, a guideline on the management of breast cancer may include a discussion on the overall effects on various final outcomes. These may include: survival, quality of life, adverse effects, and symptom management or a discussion comparing one treatment option to another. There should be evidence that these issues have been addressed." It also suggests that the potential cost implications of applying the recommendations should have been considered. In general, desirable and undesirable effects should be considered in all guidelines. Undesirable effects include harms (including the possibility of unanticipated

adverse effects), burdens (e.g. having to go to the doctor) and costs (including opportunity costs).

The GRADE Working Group suggests that explicit judgments should be made about which outcomes are critical, which ones are important but not critical, and which ones are unimportant and can be ignored. The group emphasizes that all important outcomes should be considered in making a recommendation, but only critical ones should be considered when making judgements about the overall quality of the evidence underlying a recommendation [19]. They recommend that it is important to consider costs (resource utilisation) before making a recommendation. They also suggest that studies using surrogate outcomes generally provide weaker evidence than those using outcomes that are important, and these only should be included when evidence for important outcomes is lacking. Thus, important outcomes (e.g. mortality, morbidity, quality of life) should be preferred over surrogate, indirect outcomes (e.g. cholesterol levels, lung function) that may or may not correlate with patient important outcomes.

Because the importance of different outcomes can vary dramatically and the importance attached to different outcomes may vary from culture to culture, it is important to take cultural diversity into account when deciding which outcomes are important [20-23]. Prenatal screening and genetic counseling are examples of interventions for which the importance of an outcome (abortion) varies between individuals and across cultures, because of religious beliefs or values [24,25]. End of life decisions are influenced by the roles of decision makers (clinician versus patient and family) and cultural differences [20,22]. The choice of using aspirin is related to the values and preferences of diabetic patients and patients place very different values on preventing strokes than their health care providers [1,26]. Cultural differences can be taken into account through the involvement of stakeholders from different cultures, and may require that judgments about trade-offs between the benefits and downsides of an intervention are specific for different cultures [27-29]. Values of stakeholders should be elicited and transparently described in recommendations. We offer strategies in another article of this series [2].

Ethical considerations should also be taken into account when selecting outcomes. For example, with directly observed therapy for tuberculosis, individual rights to refuse therapy (autonomy) may have to be sacrificed for the benefit of society [30]. Explicit identification of ethical consequences, and explicit judgments about trade-offs such as these, can help to ensure that appropriate judgments are made, help to resolve or clarify disagreements, and facilitate local adaptation of guidelines.

How should the importance of outcomes be ranked?

Judgments about the balance between the benefits and downsides of an intervention require judgments about the relative importance of the different outcomes, either explicitly or implicitly. Ranking outcomes by their relative importance, separated into benefits and downsides in an evidence profile [7,12] can help to focus attention on those outcomes that are considered most important, and help to resolve or clarify disagreements. Research on values and preferences should guide the ranking of outcomes, whenever possible. Guideline panels may want to search for research on the values associated with specific outcomes of interest to inform judgments about their relative importance.

If the importance of outcomes varies across cultures, ranking should be done by people in a specific setting, who can pay due consideration to local values and preferences. If evidence is lacking for an important outcome, this should be acknowledged, rather than ignoring the outcome.

Discussion

There is very limited evidence to inform decisions about how to select and rank outcomes. However, we recommend the use of systematic and transparent methods involving key stakeholders, including consumers and people from different cultures, to help ensure that all important outcomes are considered and facilitate local adaptation of guidelines. Limitations of our work include the possibility that we have missed relevant studies.

Further work

Although it is possible that there is relevant empirical research of which we are not aware, a complete systematic review of the questions addressed in this paper is unlikely to change the conclusion that there is very little research evidence in this area. Evaluations comparing different methods of identifying, selecting and ranking outcomes are needed.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

HJS prepared the first draft of this introduction. ADO and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper. The authors thank Mr. John Fox for administrative assistance with the literature review and organizing the literature database.

References

1. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, Nagpal S, Cox JL: **Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study.** *Bmj* 2001, **323(7323)**:1218-1222.
2. Schünemann HJ, Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 10. Integrating values and consumer involvement.** *Health Res Policy Syst* 2006.
3. Woolf SH, DiGuseppi CG, Atkins D, Kamerow DB: **Developing evidence-based clinical practice guidelines: lessons learned by the US Preventive Services Task Force.** *Annu Rev Public Health* 1996, **17**:511-538.
4. McKibbon A, Hunt D, Richardson SW, al. : **Finding the evidence.** In *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* Edited by: Guyatt G RD. Chicago , American Medical Association Press; 2002:16.
5. National Health and Medical Research C: **Handbook series on preparing clinical practice guidelines. February 2000. Australia.** 2000.
6. National Health and Medical Research C: **NHMRC additional levels of evidence and grades for recommendations for developers of guidelines PILOT PROGRAM 2005.** .
7. **Scottish Intercollegiate Guideline Network. SIGN 50: A guideline developers' handbook. Section 5.** :(accessed on March 17, 2006) [<http://www.sign.ac.uk/guidelines/fulltext/50/section5.html>].
8. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D: **Current methods of the U.S. Preventive Services Task Force: A review of the process.** *American Journal of Preventive Medicine* 2001, **20(3, Supplement 1)**:21-35.
9. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006.
10. **Agency for Healthcare Research and Quality** [<http://www.guidelines.gov>]
11. **Guidelines International Network** [<http://www.g-i-n.net>]
12. **National Institutes for Health and Clinical Excellence** [http://www.nice.org.uk/pdf/GDM_Allchapters_0305.pdf]
13. Kelson MC: **The NICE Patient Involvement Unit.** *Evidence-Based Healthcare & Public Health* 2005, **9**:304-307.
14. Owens DK, Nease RF Jr.: **Development of outcome-based practice guidelines: a method for structuring problems and synthesizing evidence.** *Jt Comm J Qual Improv* 1993, **19(7)**:248-263.
15. 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.** *Ann Intern Med* 2005, **142(12, Part 2)**:1056-1065.
16. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A: **Applying systematic review methods to studies of people's views: an example from public health research.** *J Epidemiol Community Health* 2004, **58(9)**:794-800.
17. Watts T, Merrell J, Murphy F, Williams A: **Breast health information needs of women from minority ethnic groups.** *J Adv Nurs* 2004, **47(5)**:526-535.
18. **AGREE Collaboration (www.agreetrust.org). 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.

19. 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, Schunemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr., Zaza S: **Grading quality of evidence and strength of recommendations.** *Bmj* 2004, **328(7454)**:1490.
20. Klessig J: **The effect of values and culture on life-support decisions.** *West J Med* 1992, **157(3)**:316-322.
21. Blackhall LJ, Murphy ST, Frank G, Michel V, Azen S: **Ethnicity and attitudes toward patient autonomy.** *Jama* 1995, **274(10)**:820-825.
22. Ruhnke GV, Wilson SR, Akamatsu T, Kinoue T, Takashima Y, Goldstein MK, Koenig BA, Hornberger JC, Raffin TA: **Ethical decision making and patient autonomy: a comparison of physicians and patients in Japan and the United States.** *Chest* 2000, **118(4)**:1172-1182.
23. Weber E Hsee C: **Culture and Individual Judgment and Decision Making.** *Applied Psychology An International Review* 2000, **49**:32-61.
24. Hunt LM, de Voogd KB, Castaneda H: **The routine and the traumatic in prenatal genetic diagnosis: does clinical information inform patient decision-making?** *Patient Educ Couns* 2005, **56(3)**:302-312.
25. Fischer RL, Schaeffer K, Hunter RL: **Attitudes of obstetrics and gynecology residents toward abortion participation: a Philadelphia area survey.** *Contraception* 2005, **72(3)**:200-205.
26. Montori VM, Bryant SC, O'Connor AM, Jorgensen NW, Walsh EE, Smith SA: **Decisional attributes of patients with diabetes: the aspirin choice.** *Diabetes Care* 2003, **26(10)**:2804-2809.
27. Duffy SA, Jackson FC, Schim SM, Ronis DL, Fowler KE: **Racial/ethnic preferences, sex preferences, and perceived discrimination related to end-of-life care.** *J Am Geriatr Soc* 2006, **54(1)**:150-157.
28. Suurmond J, Seeleman C: **Shared decision-making in an intercultural context. Barriers in the interaction between physicians and immigrant patients.** *Patient Educ Couns* 2006, **60(2)**:253-259.
29. Rebagliato M, Cuttini M, Broggin L, Berbik I, de Vonderweid U, Hansen G, Kaminski M, Kollee LA, Kucinskas A, Lenoir S, Levin A, Persson J, Reid M, Saracci R: **Neonatal end-of-life decision making: Physicians' attitudes and relationship with self-reported practices in 10 European countries.** *Jama* 2000, **284(19)**:2451-2459.
30. Verma G, Upshur RE, Rea E, Benatar SR: **Critical reflections on evidence, ethics and effectiveness in the management of tuberculosis: public health and global perspectives.** *BMC Med Ethics* 2004, **5**:E2.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 7. Deciding what evidence to include

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim³

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway, ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ³Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 01 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:19 doi:10.1186/1478-4505-4-19

Accepted: 01 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/19>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the seventh of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on what constitutes "evidence" in guidelines and recommendations.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key question and answers: We found several systematic reviews that compared the findings of observational studies with randomised trials, a systematic review of methods for evaluating bias in non-randomised trials and several descriptive studies of methods used in systematic reviews of population interventions and harmful effects.

What types of evidence should be used to address different types of questions?

- The most important type of evidence for informing global recommendations is evidence of the effects of the options (interventions or actions) that are considered in a recommendation. This evidence is essential, but not sufficient for making recommendations about what to do. Other types of required evidence are largely context specific.
- The study designs to be included in a review should be dictated by the interventions and outcomes being considered. A decision about how broad a range of study designs to consider should be made in relationship to the characteristics of the interventions being considered, what evidence is available, and the time and resources available.
- There is uncertainty regarding what study designs to include for some specific types of questions, particularly for questions regarding population interventions, harmful effects and interventions where there is only limited human evidence.
- Decisions about the range of study designs to include should be made explicitly.
- Great caution should be taken to avoid confusing a lack of evidence with evidence of no effect, and to acknowledge uncertainty.
- Expert opinion is not a type of study design and should not be used as evidence. The evidence (experience or observations) that is the basis of expert opinions should be identified and appraised in a systematic and transparent way.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the seventh of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Recommendations about health care and about interventions or actions that affect health, such as social or environmental interventions, can be informed by a wide range of evidence including randomised trials, non-randomised comparative studies, descriptive studies, qualitative research, animal studies and laboratory studies. Discussions of evidence-informed policy and practice can generate debates regarding what constitutes 'evidence' [1]. A common understanding of evidence is that "evidence concerns facts (actual or asserted) intended for use in support of a conclusion" [1]. A fact, in turn, is something known by experience or observation. An important implication of this understanding of evidence is that evidence is used to support a conclusion; it is not the same as the conclusion. Evidence alone does not make decisions.

This understanding of what evidence is has several implications. Firstly, expert opinion is more than evidence. It combines facts, interpretation of those facts, and conclusions. There is evidence behind expert opinions. Expert opinion should be used appropriately by identifying the facts (experience or observations) that are the basis of the opinions and appraising the extent to which the facts support the conclusions [2].

Secondly, not all evidence is equally convincing. How convincing evidence is (for effects) should be based on criteria such as: What sort of observations? How well were they done? How consistent are they? How directly relevant are they? How many are there? How strong is an association?

Thirdly, judgements about how much confidence to place in different types of evidence (the 'quality' of the evidence) are made either implicitly or explicitly. It is better to make these judgements systematically and explicitly to help protect against errors, resolve disagreements, facilitate critical appraisal, and communicate information. This, in turn, requires explicit decisions about what types of evidence to consider at all.

Fourthly, all evidence is context sensitive, since observations are made in a specific context. A judgement always needs to be made about their applicability beyond that context. It is best to make judgements about applicability

systematically and explicitly, for the same reasons that it is best to make judgements about the quality of the evidence systematically and explicitly.

Fifthly, global evidence (i.e. the best evidence from around the world) is the best starting point for judgements about effects, likely modifying factors, and (sometimes at least) resource utilisation. This argument is based on the understanding that all evidence is context sensitive to some extent and, therefore, indirect to some extent. Decisions based on a subset of observations are more prone to random errors [3], and judgements about whether to base a conclusion on a subset of observations are better informed if the overall observations (all of the relevant global evidence) are known [4].

Sixthly, local evidence (from the specific setting in which decisions and actions will be taken) is needed for most other judgements about what to do, including: the presence of modifying factors in specific settings, need (prevalence, baseline risk or status), values, costs and the availability of resources.

Recognising the need for both global evidence (of effects) and local evidence, it is important to be cautious about developing global recommendations. Nonetheless, global recommendations are valuable when different local conditions are not likely to lead to different decisions. When different conditions are likely to lead to different decisions, global frameworks for decisions are still important. These can reduce unnecessary duplication of efforts. They are particularly important to support low and middle-income countries, with limited resources to systematically develop guidelines, to make context specific decisions by providing the global evidence, a framework for decisions, and practical advice for incorporating local evidence.

WHO's focus is on global recommendations and supporting its member states to make well-informed decisions. The primary question that needs to be addressed in this context is:

- What types of study designs should be used to address different types of questions about the effects of the different options that are considered when making a recommendation?

We therefore have focused this review on questions about effects, recognising that there are parallel questions regarding what types of study designs should be used to address other questions. In addressing this question we have focused on the validity of different study designs, assuming that questions about the applicability of the results of studies to the specific questions of interest will be similar across different study designs. However, it is

important to recognise that decisions about what study designs to include may also be influenced by the extent to which relevant studies are available that have used study designs that are most likely to provide valid results. That is, there may sometimes be a trade-off between including studies that are more likely to be valid and ones that are more likely to be directly relevant.

What WHO is doing now

The Guidelines for WHO Guidelines (GWG) state: "It is recommended that [a] systematic review be undertaken (<http://hiru.mcmaster.ca/cochrane/cochrane/hbook.htm>) After the studies have been identified and critically appraised, and the evidence synthesised, evidence should be graded. All evidence, including that on safety, should be clearly laid out in an evidence table. Meta-analysis should be done when the data permit. The final results should be presented in a balance sheet" [5]. The GWG do not address the choice of study designs for different types of questions. In practice it is difficult to know what study designs are considered relevant for different types of WHO recommendations since few WHO guidelines have adhered to the GWG, few have included a systematic review, and many do not include references [6,7].

What other organisations are doing

The U.S. Preventive Services Task Force has the following approach to determining what evidence is admissible:

The topic team determines the bibliographic databases to be searched and the specific inclusion and exclusion criteria (i.e., admissible evidence) for the literature on each key question. Such criteria typically include study design, population studied, year of study, outcomes assessed, and length of follow-up. Topic teams specify criteria on a topic-by-topic basis rather than adhering to generic criteria. If high-quality evidence is available, the topic teams may exclude lower-quality studies. Conversely, if higher-quality evidence is lacking, the teams may examine lower-quality evidence.

If a search finds a well-performed systematic review that directly addresses the literature on a key question through a given date, the topic team may use this review to capture the literature for those dates. The team can then restrict its own search to dates not covered by the existing systematic review.

The topic team documents these strategies for sharpening focus – the analytic framework, key questions, and criteria for admissible evidence – in an initial work plan. This work plan is presented to the Task Force at its first meeting after the topic has been assigned, allowing the Task Force the opportunity to modify the direction and scope of the review, as needed [8].

This approach is consistent with other guidance for systematic reviews, such as those of the Cochrane Health Pro-

motion and Public Health Task Force, which recommends that: "The study designs to be included in a public health review should be dictated by the interventions being reviewed (methodological appropriateness), and not vice versa" [9]. There is also general, although not unanimous, agreement that the inclusion criteria for a systematic review should specify the study designs that are acceptable for a specific question [10]. However, there are important differences in both guidance and practice with respect to "how low" reviewers should go in deciding what evidence to include [11]. This question is particularly relevant for questions about the effects of population interventions (public health, health promotion, health systems and social interventions) and for evidence of harmful effects [10-20].

The Cochrane Handbook for Systematic Reviews of Interventions takes a relatively cautious approach: "The more restrictive authors are in matching questions to particular aspects of design, the less likely they are to find data specific to the restricted question. However, reviewing studies that are unlikely to provide reliable data with which to answer the question is a poor use of time and can result in misleading conclusions." [21] Because Cochrane reviews address questions about the effects of health care, they focus primarily on randomised trials. The Handbook suggests being cautious of including non-randomised studies because of the risk of biased results; the additional work required to identify and appraise non-randomised studies and keep a review up-to-date; and the risk of publication bias. It concludes: "While attention to the risk of bias should guide decisions about what types of study designs to include in a review, individual authors and Collaborative Review Groups must decide what types of studies are best suited to specific questions."

Within the Cochrane Collaboration, several groups have recommended inclusion of a broader range of study designs for health systems and public health interventions and for assessing harmful effects of clinical interventions. The Cochrane Effective Practice and Organisation of Care Group (EPOC) argues that: *While cluster randomised trials are the most robust design for quality improvement strategies, some strategies may not be amenable to randomisation – for example, mass media campaigns. Under these circumstances, reviewers may choose to include other designs including quasi-experimental designs. If a review includes quasi-experimental studies – for example, interrupted time series designs for evaluating mass media campaigns, the reviewers need to recognise the weaknesses of such designs and be cautious of over-interpreting the results of such studies. Within EPOC, reviewers can include randomised trials, controlled before and after studies, and interrupted time series [17].*

The Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Taskforce suggests including a still broader range of study designs: "A wide

variety of study designs may be used in the evaluation of public health activities, ranging from randomized controlled trials (RCTs) to case studies, with no single method being able to answer all relevant questions about the effectiveness of all public health interventions." [9]

The Cochrane Adverse Effects Subgroup identifies three possible approaches for incorporating adverse effect data in a review and summarises the advantages and disadvantages of each of these approaches as summarised in Table 1[18,19].

The U.K. NHS Centre for Reviews and Dissemination provides the following guidance: "The inclusion criterion specifying the type of study design stems from the desire to base reviews on the highest quality evidence. There are several areas of health care which have not been evaluated with methodologically sound studies. In this situation, studies of methodologically lower quality may have to be included. Here it is important to note that the preference for one or another study design should depend on the nature of questions raised in the review. Inevitably the decisions regarding inclusion based on study design will also depend on the availability of suitable study designs in the literature." [22]

We are not aware of any specific guidance for what study designs to include for non-human studies, although some recommendations rely on animal and in vitro studies. For example, treatment recommendations for emerging dis-

eases, such as SARS or avian influenza (H5N1), for which case reports may be the only human studies that are available, may be based on a combination of indirect human evidence (from the treatment of other similar diseases), case reports, animal studies and in vitro studies. In general, the same principles that apply to human studies can be applied to animal and in vitro studies [23].

The Guide to Community Preventive Services uses data from comparative studies – those that compare outcomes among a group exposed to the intervention versus outcomes in a concurrent or historical group that was not exposed or was less exposed – to answer questions about whether interventions are effective [24]. All comparative studies are included in its reviews, assessed for their design suitability and threats to internal and external validity, and assessed for potential effects of study design and execution on results.

The Campbell Collaboration does not provide specific guidance on what study designs should be used to address different types of questions related to the effects of interventions in the social, behavioral and educational arenas [25].

Methods

The methods used to prepare this review are described in the introduction to this series [26]. Briefly, the key questions addressed in this paper were vetted amongst the

Table 1: Pros and cons of different approaches for incorporating adverse effect data in a systematic review*

Method	Look in the trials/studies included in the systematic review of benefit.	Look in all retrieved trials/studies of that intervention, even in those excluded from the analysis of benefit	Look for studies that specifically evaluate adverse effects of the intervention
Protocol	Should usually be the minimum recommendation	Studies rejected from analysis of benefit (e.g. because beneficial outcomes are measured in a different way, which cannot be combined with other studies), may be included to allow adverse effect data collection. Two sets of inclusion criteria will be needed – for benefit, and for adverse effects	Design separate strategy to identify studies that report adverse effects, including those that do not look at beneficial effects.
Pros	Less demanding on time and resources Does not require new literature search strategy	More comprehensive than just looking at included trials Can potentially cover a more representative group of patients	Might amount to a separate review nested within a traditional Cochrane review Most comprehensive
Cons	Data may be very limited and biased towards common, short-term harms	Relatively time consuming as full-text articles of all potentially relevant studies need checking Data may be limited to well-recognized and commonly seen adverse effects. Benefit and harm cannot be compared directly as the data come from different sources	May be able to evaluate rare, or long-term, or previously unrecognized adverse effects Time and resource intensive Special techniques required in synthesizing data from a diverse range of sources Increased quantity of data but greater risk of biased and poor quality data Benefit and harm cannot be compared directly as the data come from different sources.

*Copied from reference [18].

authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register, the US National Guideline Clearinghouse, and the Guidelines International Network for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review we searched PubMed using (clinical practice guidelines or public health guidelines or systematic reviews) and (study designs) and related articles for references. We searched the Cochrane Methodology Register using the key word study design, and we checked the reference lists of the reports that we retrieved. The searches were conducted in February and March 2006.

Findings

We found several systematic reviews that compared the findings of observational studies with randomised trials [27-33], and a systematic review of methods for evaluating bias in non-randomised trials [34]. We also found several descriptive studies of methods used in systematic reviews of population interventions and harmful effects.

Systematic reviews of the results of randomised trials compared with observational studies have differed in the methods they have used, and, to some extent, in their conclusions, but have generally found that it is not possible to predict differences in the size, or even the direction, of estimates of treatment effects for the same intervention when it is generated in randomized and non-randomized studies. However, especially in the more recent reports [30-33], there is the suggestion that these disparities decrease when investigators have controlled for known confounders (between risk/responsiveness and treatment).

The review of methods for evaluating bias in non-randomised trials found six tools that were thought to be suitable for use in systematic reviews [34]. Their review of 511 systematic reviews that included non-randomised studies found that only 169 (33%) assessed study quality. A more recent survey of methods used in systematic reviews of adverse effects found that although more than three quarters (185/243) reviews sought to include data from sources other than randomised controlled trials, fewer than half (106/256) assessed the quality of the studies that were included [35].

A study that considered the potential of randomised trials to provide evidence on specific harms found that of 1727

Cochrane reviews, only 138 included evidence on ≥ 4000 subjects. Of these only 25 (18%) had eligible data on adverse events, while 77 had no harms data, and 36 had data on harms that were non-specific or pertained to < 4000 subjects [17]. Thus, while systematic reviews of randomised trials can provide useful information on adverse effects of clinical interventions, the reporting of adverse effects in both randomised trials and systematic reviews needs to be improved.

Descriptive reports of reviews of harmful effects have found that a significant investment of effort failed to yield significant new information [18,19,36]. Authors of reviews of social interventions, on the other hand, have argued that restricting the study designs that are included in a review may reduce the value of the review and reinforce the "inverse evidence law" whereby the least is known about the effects of interventions most likely to influence whole populations. However, this argument relates more to the importance of mapping out the available evidence than to producing reliable estimates of the effects of interventions [11].

Discussion

While there is broad agreement that the study designs to be included in a review should be dictated by the interventions being reviewed, there is uncertainty regarding what study designs to include for some specific types of questions. For any question, as the range of study designs that are included is broadened, an increasing amount of work is required to derive decreasingly reliable estimates of the effects of interventions. A decision about how broad a range of study designs to consider must be made in relationship to the characteristics of the interventions, what evidence is available, and the time and resources available.

For any question there is a cut-off point beyond which broadening the types of studies that are considered requires a substantial investment of effort that will not yield additional information that will inform decisions in a meaningful way. In many cases, it is likely to be prudent to acknowledge the limits of what is known from a restricted range of study designs, rather than to invest additional resources that are unlikely to do more than confirm the limits of what is known. Whatever decision is taken about the range of study designs to include should be made explicit, and great caution should be taken to avoid confusing a lack of evidence with evidence of no effect.

Further work

There is a rapidly growing number of reviews and studies comparing the results of different study designs. High priority should be given to generating and periodically

updating a common data set of studies to update and reconcile different conclusions among these reviews. Priority should also be given to broadening the scope of these comparisons to include a wider range of questions and a wider range of study designs, including animal and laboratory studies. Additional studies, and systematic reviews of studies, that more rigorously assess the added cost and value of including broader ranges of study designs would help to inform decisions about when it is likely to be important and worthwhile to use more diverse types of study designs. There is a need to develop more detailed guidance regarding decisions for which study designs to include for different types of questions for incorporation in the Guidelines for WHO Guidelines. This guidance, which is particularly needed for harms and interventions targeted at populations, should be based on both empirical evidence and conceptual arguments.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Lomas J, Culyer T, McCutcheon C, McAuley L, Law S: **Conceptualizing and Combining evidence for health system guidance.** Ottawa: Canadian Health Services Research Foundation; 2005.
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 9. Grading evidence and recommendations.** *Health Res Policy Syst* 2006.
- Counsell CE, Clarke MJ, Slattery J, Sandercock PA: **The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis?** *BMJ* 1994, **309**:1677-81.
- Oxman AD, Guyatt G: **When to believe a subgroup analysis.** In *Users' Guide to the Medical Literature. A Manual for Evidence-Based Clinical Practice* Edited by: Guyatt G, Rennie D. Chicago: AMA Press; 2002:553-65.
- Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines.** Geneva: World Health Organisation; 2003. (EIP/GPE/EQC/2003.1)
- Panisset U: **A review of WHO recommendations published in 2005.**
- Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.**
- Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, for the Methods Word Group, third U.S. Preventive Services Task Force: **Current methods of the U.S. Preventive Services Task Force: a review of the process.** *Am J Prev Med* 2001, **20(3S)**:21-35.
- Jackson N, Waters E, for the Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce: **Criteria for the systematic review of health promotion and public health interventions.** *Health Promotion International* 2005, **20**:367-74.
- Sheldon T: **Making evidence synthesis more useful for management and policy-making.** *J Health Serv Res Policy* 2005, **10(Suppl 1)**:1-5.
- Ogilvie D, Egan M, Hamilton V, Petticrew : **Systematic reviews of health effects of social interventions: 2. best available evidence: how low should you go?** *J Epidemiol Community Health* 2005, **59**:886-92.
- Nutbeam D: **The challenge to provide 'evidence' in health promotion.** *Health Promotion International* 1999, **14**:99-101.
- Petticrew M: **Presumed innocent. Why we need systematic reviews of social policies.** *Am J Prev Med* 2003, **23(suppl 3)**:2-3.
- Petticrew M, Roberts H: **Evidence, hierarchies, and typologies: horses for courses.** *J Epidemiol Community Health* 2003, **57**:527-9.
- Rychetnik L, Frommer M, Hawe P, Shiell A: **Criteria for evaluating evidence on public health interventions.** *J Epidemiol Community Health* 2002, **56**:119-27.
- Kramer MS: **Randomized trials and public health interventions: time to end the scientific double standard.** *Clinics in Perinatology* 2003, **30**:351-61.
- Grimshaw J, McAuley LM, Bero LA, Grilli R, Oxman AD, Ramsay , Vale L, Zwarenstein M: **Systematic reviews of the effectiveness of quality improvement strategies and programmes.** *Qual Saf Health Care* 2003, **12**:298-303.
- Loke YK, Price D, Herxheimer A, editors: **Including adverse effects.** *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]* . Appendix 6b
- McIntosh HM, Nerys NF, Woolacott F, Bagnall AM: **Assessing harmful effects in systematic Reviews.** *BMC Medical Research Methodology* 2004, **4**:19.
- Papanikolaou PN, Ioannidis JPA: **Availability of large-scale evidence on specific harms from systematic reviews of randomized trials.** *Am J Med* 2004, **117**:582-9.
- Higgins JPT, Green S, editors: **4.2.4 What types of study designs?** *Cochrane Handbook for Systematic Reviews of Interventions [updated May 2005]* [<http://www.cochrane.org/resources/handbook/hbook.htm>]. Section 4
- Khan KS, Kleijnen J: **STAGE II. Conducting the review. PHASE 4. Selection of studies.** *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4* 2nd edition. 2001 [http://www.york.ac.uk/inst/crd/pdf/crd4_ph4.pdf].
- Festing MFV, Altman DG: **Guidelines for the Design and Statistical Analysis of Experiments Using Laboratory Animals.** *ILAR Journal* 2002, **43**:244-58.
- Briss PA, Zaza S, Pappaioanou M, et al.: **Developing an evidence-based guide to community preventive services-methods.** *Am J Prev Med* 2000, **18(1S)**:35-43.
- Campbell Systematic Reviews. Guidelines for the Preparation of Review Protocols. (Version 1.0: January 1, 2001)** [http://www.campbellcollaboration.org/c2_protocol_guidelines%20doc.pdf]
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Heinsman DT, Shadish VWR: **Assignment methods in experimentation: when do nonrandomized experiments approximate answers from randomized experiments?** *Psychological Methods* 1996, **1**:154-69.
- Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C: **Choosing between randomised and non-randomised studies: a systematic review.** *Health Technol Assessment* 1998, **2**:#13.
- Kunz R, Vist G, Oxman AD: **Randomisation to protect against selection bias in healthcare trials.** *The Cochrane Database of Methodology Reviews* 2002.

30. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS: **A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.** *Health Technol Assessment* 2000, **4(34)**.
31. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MB, Contopoulos Ioannidis DG, Lau J: **Comparison of evidence of treatment effects in randomized and nonrandomized studies.** *JAMA* 2001, **286**:821-30.
32. King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R: **Impact of participant and physician intervention preferences on randomized trials; a systematic review.** *JAMA* 2005, **293**:1089-99.
33. Vist GE, Hagen KB, Devereaux PJ, Bryant D, Kristoffersen DT, Oxman AD: **Systematic review to determine whether participation in a trial influences outcome.** *BMJ* 2005, **330**:1175.
34. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakaravitch C, Song F, et al.: **Evaluating non-randomised intervention studies.** *Health Technol Assess* 2003, **7(27)**:1-173.
35. Golder SP, Loke YK, McIntosh HM: **Room for improvement? A survey of the methods used in systematic reviews of adverse effects.** *BMC Medical Research Methodology* 2006, **6**:3.
36. Loke YK, Derry S: **Incorporating adverse effects data into reviews: how to get started [abstract].** *9th Annual Meeting for UK Contributors to the Cochrane Collaboration: Coventry 2003.*

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim¹

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 05 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:20 doi:10.1186/1478-4505-4-20

Accepted: 05 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/20>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the eighth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on the synthesis and presentation of research evidence, focusing on four key questions.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We found two reviews of instruments for critically appraising systematic reviews, several studies of the importance of using extensive searches for reviews and determining when it is important to update reviews, and consensus statements about the reporting of reviews that informed our answers to the following questions.

How should existing systematic reviews be critically appraised?

- Because preparing systematic reviews can take over a year and require capacity and resources, existing reviews should be used when possible and updated, if needed.
- Standard criteria, such as A MeaSurement Tool to Assess Reviews (AMSTAR), should be used to critically appraise existing systematic reviews, together with an assessment of the relevance of the review to the questions being asked.

When and how should WHO undertake or commission new reviews?

- Consideration should be given to undertaking or commissioning a new review whenever a relevant, up-to-date review of good quality is not available.
- When time or resources are limited it may be necessary to undertake rapid assessments. The methods that are used to do these assessments should be reported, including important limitations and uncertainties and explicit consideration of the need and urgency of undertaking a full systematic review.
- Because WHO has limited capacity for undertaking systematic reviews, reviews will often need to be commissioned when a new review is needed. Consideration should be given to establishing collaborating centres to undertake or support this work, similar to what some national organisations have done.

How should the findings of systematic reviews be summarised and presented to committees responsible for making recommendations?

- Concise summaries (evidence tables) of the best available evidence for each important outcome, including benefits, harms and costs, should be presented to the groups responsible for making recommendations. These should include an assessment of the quality of the evidence and a summary of the findings for each outcome.
- The full systematic reviews, on which the summaries are based, should also be available to both those making recommendations and users of the recommendations.

What additional information is needed to inform recommendations and how should this information be synthesised with information about effects and presented to committees?

- Additional information that is needed to inform recommendations includes factors that might modify the expected effects, need (prevalence, baseline risk or status), values (the relative importance of key outcomes), costs and the availability of resources.
 - Any assumptions that are made about values or other factors that may vary from setting to setting should be made explicit.
 - For global guidelines that are intended to inform decisions in different settings, consideration should be given to using a template to assist the synthesis of information specific to a setting with the global evidence of the effects of the relevant interventions.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the eighth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

A summary of the best available research evidence is essential, though not sufficient to inform recommendations. To reduce the risk of bias and errors that occur by chance, and to facilitate critical appraisal of syntheses of evidence, reviews should be systematic and should explicitly report the methods that were used [1]. However, systematic reviews require resources, take time, and may not always be warranted or possible. Moreover, unnecessary duplication of systematic reviews should be avoided, given the large unmet need for systematic reviews of a wide range of questions and the need to keep reviews up-to-date [2,3].

The first step in considering the needs for systematic reviews to inform recommendations is to critically appraise existing reviews to determine if they provide an adequate summary of the relevant evidence that is needed, particularly evidence of the effects of the different options (interventions) that are being considered. If they do not, consideration must then be given to whether a new review should be undertaken and how best to obtain a new review. Once an adequate summary of the evidence is available, consideration must be given to how best to

present that information to the group of people who will consider that evidence, together with other evidence and judgements, to develop recommendations. In addition, consideration needs to be given to the additional information that is needed and how that should be summarised and presented.

In this paper we address the following questions:

- How should existing systematic reviews be critically appraised and used?
- When and how should WHO undertake or commission new reviews?
- How should the findings of systematic reviews be summarised and presented to committees responsible for making recommendations?
- What additional information is needed to inform recommendations and how should this information be synthesised with information about effects and presented to committees?

Related questions on priority setting for guidelines or recommendations and reporting of guidelines are addressed in other papers in this series [4,5].

What WHO is doing now

Of 62 WHO documents that were indexed as guidelines in 2005, only two reported a systematic review and less than 40% included references [6]. Although it is possible that systematic reviews are being used and this is not being

reported, this is unlikely. With some notable exceptions, for the most part recommendations are currently being made without adequate use of existing systematic reviews and systematic reviews are rarely being undertaken or commissioned by WHO committees that make recommendations. However, the situation may be somewhat better than what is reported in published guidelines. When asked about the use of evidence of effects specifically in an interview study [7], many departments reported using background documents. These were reported to have been prepared in a variety of ways, including as unpublished working papers, documents similar to those used by the Scottish Intercollegiate Guidelines Network (SIGN), and documents prepared by the participating experts. Only two departments reported using systematic reviews specifically, while several others reported using systematic reviews along with a range of other documents. Others reported leaving the use of evidence up to the experts, a lack of documentation, evidence of effects not being relevant for some recommendations, and using a mixture of "epidemiological data, trial data, opinions based on logical reasoning (common sense) and clinical experience."

No departments reported using concise summaries of findings or "balance sheets" for the most important outcomes (benefits, harms and costs) for the options that were considered. WHO groups that develop recommendations are, for the most part, composed of experts in a particular content area and not supported by experts in particular methodological areas (e.g. systematic reviews) or by staff with particular technical skills (e.g. information retrieval). Relatively little attention appears to have been given to how best to help member states adapt global recommendations, taking account of local needs, values, resources and conditions.

What other organisations are doing

In contrast, in a survey of 101 organisations that produce clinical practice guidelines 95% of the 58 respondents reported that they provide guideline panels with systematic reviews [8]. In another survey of 18 prominent organisations that develop clinical practice guidelines, all but one reported using systematic reviews [9].

The UK National Centre for Health and Clinical Excellence (NICE), for example routinely undertakes systematic reviews to inform its guideline panels [10]. NICE has seven professionally led National Collaborating Centres to manage the development of clinical guidelines [11]. Each Centre has a range of skills and abilities, including systematic reviewing. The Centres are responsible for identifying the best and most relevant evidence available. They write the first consultation draft of a guideline over a

period of 12 to 18 months. NICE reviews are available in the full version of its guidelines.

Other organisations that produce guidelines sometimes use existing systematic reviews, sometimes prepare their own systematic reviews, and sometimes commission reviews. The U.S. Preventive Services Task Force, for example, commissions systematic reviews from Evidence-based Practice Centers (EPCs) for updates of its guidelines [12]. The Agency for Healthcare Research and Quality (AHRQ) has contracts with 13 EPCs from which it commissions systematic reviews. AHRQ does not produce guidelines, but stakeholder organisations that request the reviews may produce guidelines. Other health technology assessment (HTA) agencies, which may or may not produce guidelines, have staff that undertake reviews, convene expert groups that undertake reviews together with support from staff, or commission systematic reviews [13].

Systematic reviews of the effects of interventions are a major focus for most organisations that develop guidelines. Because most organisations develop recommendations for a specific country or setting, they are able to take into account additional information relevant to the specific context for which the recommendations are intended, including factors that might affect the applicability of the evidence in specific settings, need (prevalence, baseline risk or status), values, costs and the availability of resources.

Methods

The methods used to prepare this review are described in the introduction to this series [14]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register [15], the US National Guideline Clearinghouse [16], and the Guidelines International Network [17]) for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review we knew of two previous systematic reviews of instruments for critically appraising systematic reviews through personal contacts [18,19], and studies of how to present the results of systematic reviews to policy makers [20], the general public [21], and users of Cochrane reviews [22]. We used these studies and their reference lists to identify related articles in PubMed. We searched the Cochrane Methodology Register using the key word 'Presentation of reviews: General' and we

checked the reference lists of the reports that we retrieved. We searched for literature on priority setting for guidelines and health technology assessments for another report [4]. In addition, we searched broadly for literature on commissioning systematic reviews in PubMed (commissioning systematic reviews) and using Google ("commissioning systematic reviews" and "updating systematic reviews") and in the Cochrane Methodology Register using the terms 'commissioning' and 'updating systematic reviews'. The searches were conducted in March 2006.

Findings

How should existing systematic reviews be critically appraised?

The first of two reviews of different instruments for critically appraising systematic reviews found 20 systems concerned with the appraisal of systematic reviews or meta-analyses, including one scale, 10 checklists, and nine guidance documents [18]. The authors identified seven key domains that they considered important to appraise: study question, search strategy, inclusion and exclusion criteria, data abstraction, study quality, data synthesis and analysis, and funding or sponsorship. One checklist fully addressed all seven domains [23]. A second checklist also addressed all seven domains but merited only a "Partial" score for study question and study quality [24]. Two additional checklists and the one scale addressed six of the seven domains [25-27]. These latter two checklists excluded funding; the scale omitted data abstraction and had a "Partial" score for search strategy. The authors concluded that based on coverage of the seven domains that they considered key, these five systems (four checklists and one scale) represented "best practice" (i.e. were the best available instruments) for appraising systematic reviews. Although they considered other aspects of the systems, such as the methods used to select items and inter-rater reliability, they did not take these factors into consideration in their selection of these five systems, nor did they consider the suitability of the different systems for specific purposes.

The second review used a detailed process to evaluate and select a system and expanded the work by AHRQ up until the year 2005 [19]. They identified approximately 240 quality assessment instruments for systematic reviews, randomized controlled trials and observational studies as well as nearly 50 evidence grading systems. The instruments and systems identified were evaluated by type of study using the AHRQ evaluation grids from the first review, and considering descriptive items for most potential instruments and systems. The highest scoring instruments and systems from each grid represented the proposed selections. The proposed selections were then sent to the same experts that were contacted to review and provide comment during the initial expert consultation.

Based on the second expert consultation, the AMSTAR 2005 was selected as the best instrument for appraising systematic reviews (Table 1). A description of the rationale for selecting that instrument is not available.

When and how should WHO undertake or commission new reviews?

There is wide agreement that guidelines should be informed by systematic reviews of the best available evidence among organisations that develop clinical practice guidelines and, increasingly, among organisations that develop guidance for population interventions (public health, health promotion, health systems and social interventions) [8,9,28-34]. Thus, priorities for systematic reviews are set, to some extent, when a decision is first made to develop recommendations. We reviewed the methodological literature relevant to priority setting for guidelines and health technology assessments, which overlaps largely with priority setting for systematic reviews, in our review on setting priorities for developing recommendations [4]. Additional questions related to undertaking or commissioning new reviews include: If there is a systematic review is it of good enough quality and recent enough that a new review is unlikely to be needed? Are there sufficient time and resources to commission or undertake a new review, if one is needed? If there is time, resources and a need for a new review, what is the best approach to getting the work done?

The first of these questions can be answered by considering the criteria discussed above and the likelihood of whether new research is likely to have been completed. Under some circumstances, it may not be warranted or possible to undertake or commission a systematic review even if there is not a previous systematic review; for example, for emerging diseases when it is known that the available evidence is sparse and when decisions must be made urgently.

We address which evidence should be used to address different types of questions in another paper in this series [35]. As we suggest in that paper, there is a cut-off point beyond which broadening the types of studies that are included requires a substantial investment of effort that will not yield additional information that usefully informs decisions. Similarly, there is a cut-off point beyond which more extensive searches are unlikely to yield additional useful studies.

An assessment of 159 systematic reviews with comprehensive literature searches found that the importance of trials that are difficult to locate may vary, but that generally in situations where resources are limited, thorough quality assessments should take precedence over extensive literature searches and translations of articles [36,37]. Consist-

Table 1: A Measurement Tool to Assess Reviews (AMSTAR), 2005 (from COMPUS [19])**1. Was an 'a priori' design provided?**

The research question and inclusion criteria should be established before the conduct of the review.

Yes

No

Can't answer

Not applicable

2. Were there duplicate study selection and data extraction?

There should be at least two independent data extractors and the consensus procedure for disagreements should be reported.

Yes

No

Can't answer

Not applicable

3. Was a comprehensive literature search performed?

At least two electronic sources should be searched. The report must include years and databases (e.g., Central, EPOC, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

Yes

No

Can't answer

Not applicable

4. Was the status of publication (i.e., grey literature) used as an exclusion criterion?

The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status.

Yes

No

Can't answer

Not applicable

5. Was a list of studies (included and excluded) provided?

A list of included and excluded studies should be provided.

Yes

No

Can't answer

Not applicable

6. Were the characteristics of the included studies provided?

In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported.

Yes

Table 1: A Measurement Tool to Assess Reviews (AMSTAR), 2005 (from COMPUS [19]) (Continued)

No

Can't answer

Not applicable

7. Was the scientific quality of the included studies assessed and reported?

'A priori' methods of assessment should be reported (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

Yes

No

Can't answer

Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?

The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

Yes

No

Can't answer

Not applicable

9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable, to assess the homogeneity (i.e., Chi-squared test for homogeneity, I²). If heterogeneity exists, random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).

Yes

No

Can't answer

Not applicable

10. Was the likelihood of publication bias assessed?

An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot) and statistical tests (e.g., Egger regression test).

Yes

No

Can't answer

Not applicable

11. Was the conflict of interest stated?

Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

Yes

No

Can't answer

Not applicable

Source: AMSTAR 2005 (Beverley Shea, CIET, Institute of Population Health, Ottawa: personal communication, 2005 Oct)

ent with this, another assessment of Cochrane reviews found that additional database searching beyond the Cochrane Central Register of Controlled Trials (CENTRAL) retrieved only a small percentage of extra trials, and that contacting authors and manufacturers to find unpublished trials appeared to be a more effective method of obtaining additional better quality trials [38].

Similarly, a third assessment of 20 Technology Assessment Reports by NICE found that a more selective approach to database searching would suffice in most cases and would save resources, whereas searching other sources, including contact with experts and checking reference lists, appeared to be a more productive way of identifying further studies [39]. Searching additional databases beyond the Cochrane Library, MEDLINE, EMBASE and SCL, plus BIOSIS limited to meeting abstracts only, was seldom found to be effective in retrieving additional studies for inclusion in the clinical and cost-effectiveness sections of Technology Assessment Reports (apart from reviews of cancer therapies, where a search of the ASCO database was recommended).

Information retrieval for systematic reviews for public health and other non-clinical interventions may be more elusive than retrieval for reviews in clinical medicine, due to the interdisciplinary nature of the research, use of research designs other than randomised trials, and limitations of what and how the research is indexed. While it may be important to consider other databases, strategies other than database searching are likely to be important [40,41]. Moreover, database searching in public health and other non-clinical areas may require specialised skills due to technical demands of the databases to be searched, lack of standardization of the vocabulary, and the relative scarcity of rigorous evaluations [42]. Information retrieval specialists may require a broad exposure to databases, the grey literature and the terminology that is used.

Several investigators have addressed the question of when a review or guideline needs updating [37,43-47]. French and colleagues found that of a sample of 254 updated Cochrane reviews 23 (9%) had a change in conclusion [43]. Another survey of Cochrane reviews found that of 104 updated reviews in the first half of 2003, 77% included no new data or data insufficient to influence the conclusion. In 16% new data had some impact on conclusions without major change, and in only 5% new data resulted in major changes in conclusions [44].

Johnston and colleagues, on the other hand, found that an updating strategy for cancer practice guidelines found 80 pieces of new evidence over a one-year period relating to 17 of 20 guidelines [45]. On average four pieces of new evidence were found per guideline, but there was consid-

erable variation across the guidelines. Of the 80 pieces, 19 contributed to modifications of clinical recommendations in six practice guidelines, whereas the remaining evidence supported the original recommendations. In this case the updating process was resource intensive, but yielded important findings. However, it was possible to reduce the scope of the sources searched routinely to MEDLINE, the Cochrane Library and meeting proceedings. Another review of 17 guidelines published by AHRQ found that for seven guidelines new evidence and expert judgement indicated an update was needed, six were found to be in need of a minor update, three were considered still valid, and no conclusion was drawn for one [47]. The authors found that no more than 90% of the guidelines were still valid after 3.6 years and they estimated that about half the guidelines were outdated in 5.8 years. They concluded that guidelines should be reassessed every three years.

Comprehensive reviews are time-consuming. Many health technology assessment (HTA) agencies have established rapid assessment processes, particularly for new technologies [48-51]. There is no common definition of "rapid assessment" and there is variation in the scope, methods and time to complete assessments. While the concept is intuitively sound, there is little empirical evidence comparing alternative methods or comparing rapid assessments with more comprehensive methods. Milne and colleagues have described a range of HTA responses available in the UK, including 2-3 page assessments that take six weeks, rapid systematic reviews that take 8-10 weeks, technology assessment reviews that take six months, Cochrane reviews, and full HTA reports that take 3 years [52]. They identify three factors that determine the HTA response: what decision-makers want, including the time scale for decision making; the characteristics of the technology, including the importance of the uncertainty, the importance of the potential benefits, the rate of diffusion, and how much is already known from previous assessments; and the resources available for an assessment.

We did not find any evaluations of alternative methods for commissioning reviews or of comparisons between commissioning reviews and doing them in house. A survey of people preparing Cochrane reviews in Australia (with a response rate of 92/112) found that the most critical barriers to completion of a Cochrane review were lack of time (80%), lack of financial support (36%), methodological problems (23%) and problems with group dynamics (10%) [53].

How should the findings of systematic reviews be summarised and presented to committees responsible for making recommendations?

The Conference on Guideline Standardization (COGS) developed an 18-item checklist for the reporting of guidelines [29]. The checklist includes the method for synthesizing evidence (how evidence was used to create recommendations, e.g., evidence tables, meta-analysis, decision analysis) and the recommendation grading criteria (the criteria used to rate the quality of evidence that supports the recommendations and the system for describing the strength of the recommendations).

The GRADE Working Group recommends the use of evidence profiles including detailed descriptions of the judgements used to assess the quality of evidence for each important outcome and a summary of the findings for each important outcome [54,55]. More recently the Cochrane Collaboration has developed summary of findings tables, based in part on GRADE evidence profiles [22,56].

All of these methods of presenting evidence to decision makers are based on consultations informed by evidence, such as comparisons of different ways of presenting evidence. We did not find comparisons of different ways of presenting evidence to groups developing recommendations.

In addition to summaries of the main findings, such as evidence profiles, the full systematic reviews should be available to both those making recommendations and to users of the recommendations [29]. These full systematic reviews should adhere to standards such as those recommended in the QUOROM statement [57].

What additional information is needed to inform recommendations and how should this information be synthesised with information about effects and presented to committees?

Although there are a number of descriptive papers and guidelines for what additional information is needed in addition to systematic reviews of the effects of the options that are being considered, we did not find comparisons of alternative ways of synthesising this information and presenting it to groups making recommendations. As discussed in another article in this series [58], additional information that needs to be considered in a recommendation includes factors that might modify the expected effects, need (prevalence, baseline risk or status), values [59], costs and the availability of resources.

Methods of integrating this additional information and judgements include formal and informal consensus methods [60,61], decision analyses, and economic analyses

[62,63]. Because factors such as modifying factors, needs and the availability of resources can vary greatly from setting to setting, methods for incorporating this information in global guidelines are particularly challenging. We did not find any evaluations of methods for addressing these challenges.

Discussion

There is broad agreement on the need for systematic reviews to inform recommendations and on criteria for critically appraising systematic reviews. Several criteria have been identified that need to be considered when deciding whether a new systematic review is needed, including the needs of decision makers, the nature of the problem and the relevant interventions, and the availability of resources.

The available evidence suggests that, generally, in situations where time or resources are limited, thorough quality assessments should likely take precedence over extensive literature searches. When a full systematic review is not undertaken, for example because of the need for a rapid response, explicit consideration should be given to the need and urgency of undertaking a full systematic review and putting in place appropriate mechanisms for timely updating of the recommendations.

The frequency with which reviews or guidelines need to be updated is likely to vary, but as a rough rule of thumb, based in part on a study of clinical practice guidelines, the need for updating should be considered routinely after three years and more often for areas that are developing rapidly.

Further work

Both the Agency for Healthcare Research and Quality and the Canadian Coordinating Office for Health Technology Assessment have funded projects on updating systematic reviews [64,65]. These reports should help to fill in some of the gaps in this review regarding when and how to undertake or commission an update of a review. Further work is needed on several of the other questions asked in this review, including evaluation of methods for rapid assessments, how best to present evidence to groups making recommendations and, importantly for WHO, how best to take into consideration information that varies from setting to setting when making global recommendations.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS

are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Mulrow CD: **Rationale for systematic reviews.** *BMJ* 1994, **309**:597-9.
- Mallett S, Clarke M: **How many Cochrane reviews are needed to cover existing evidence on the effects of health care interventions?** *ACP J Club* 2003, **139**(1):A11.
- Chinnock P, Siegfried N, Clarke M: **Is evidence-based medicine relevant to the developing world? Systematic reviews have yet to achieve their potential as a resource for practitioners in developing countries.** *Evid Based Complement Alternat Med* 2005, **2**:321-4.
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 2. Priority setting.** *Health Res Policy Syst* 2006.
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 14. Reporting guidelines.** *Health Res Policy Syst* 2006.
- Panisett U: **A review of WHO recommendations published in 2005.**
- Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.**
- Moynihan R, Oxman AD, Lavis JN, Paulsen E: **Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier.** In *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.
- Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15**:31-45.
- Reviewing and grading the evidence** NICE: *Guideline Development Methods. National Institute for Clinical Excellence, February 2004 (updated March 2005)* [http://www.nice.org.uk/pdf/GDM_Chapter7_0305.pdf].
- National Institute for Clinical Excellence: *The Guideline Development Process: An Overview for Stakeholders, the Public and the NHS 2004* [http://www.nice.org.uk/pdf/GDP_An_Overview_for_Stakeholders_the_Public_and_the_NHS.pdf].
- Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, for the Methods Word Group, third U.S. Preventive Services Task Force: **Current methods of the U.S. Preventive Services Task Force: a review of the process.** *Am J Prev Med* 2001, **20**(3S21-35) [<http://www.elsevier.com/locate/ajpmonline>].
- Glanville J, Sowden AJ: **Identification of the need for a review.** [http://www.york.ac.uk/inst/crd/pdf/crd4_ph0.pdf].
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- National Guideline Clearinghouse** [<http://www.guidelines.gov>]
- Guidelines International Network (GIN)** [<http://www.g-i-n.net>]
- West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al.: **Systems to rate the strength of scientific evidence [Evidence report/technology assessment no 47].** 2002 [<http://www.thecre.com/pdf/ahrq-system-strength.pdf>]. Rockville (MD): Agency for Healthcare Research and Quality; AHRQ Publication No 02-E016
- Proposed Evaluation Tools for COMPUS** [<https://www.ccohta.ca/compus/>]
https://www.ccohta.ca/compus/pdf/COMPUS_Evaluation_Methodology_draft_e.pdf.
Ottawa: Canadian Coordinating Office for Health Technology Assessment November 29, 2005.
- Lavis JN, Davies HTO, Oxman AD, Denis JL, Golden-Biddle K, Ferlie E: **Towards systematic reviews that inform healthcare management and policymaking.** *Journal of Health Services Research and Policy* 2005, **10**(Suppl 1):35-48.
- Glenton C, Underland V, Kho M, Oxman AD: **Translating evidence into patient information – challenges and possible solutions when extracting and presenting the results of systematic reviews.** *J Clin Epidemiol* in press.
- Glaziovu P, Oxman AD, Higgins J: **Summary of Findings Tables within Cochrane Reviews: Draft Specification for RevMan 5.0. December 2004.** *Obtaining a consensus on the content and methods of a Summary of Findings table for Cochrane Reviews. Report to the Cochrane Collaboration Steering Group* 2005. (unpublished reports available from the author)
- Sacks HS, Reitman D, Pagano D, Kupelnick B: **Metaanalysis: an update.** *Mt Sinai J Med* 1996, **63**:216-24.
- Auperin A, Pignon JP, Poynard T: **Review article: critical review of meta-analyses of randomized clinical trials in hepatogastroenterology.** *Alimentary Pharmacol Ther* 1997, **11**:215-25.
- Irwig L, Tosteson AN, Gatsonis C, et al.: **Guidelines for metaanalyses evaluating diagnostic tests.** *Ann Intern Med* 1994, **120**:667-76.
- Khan KS, Ter Riet G, Glanville J, Sowden AJ, Kleijnen J: **Undertaking Systematic Reviews of Research on Effectiveness.** CRD's Guidance for Carrying Out or Commissioning Reviews: York, England: University of York, NHS Centre for Reviews and Dissemination; 2000.
- Stroup DF, Berlin JA, Morton SC, et al.: **Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group.** *JAMA* 2000, **283**:2008-12.
- The 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.
- 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**:493-8.
- Briss PA, Zaza S, Pappaioanou M, et al.: **Developing an evidence-based guide to community preventive services-methods.** *Am J Prev Med* 2000, **18**(15):35-43.
- Dobbins M, Thomas H, O'Brien MA, Duggan M: **Use of systematic reviews in the development of new provincial public health policies in Ontario.** *Intl J Technol Assess Health Care* 2004, **20**:399-404.
- McMichael C, Waters E, Volmink J: **Evidence-based public health: what does it offer developing countries?** *J Public Health* 2005, **27**:215-21.
- Chan KS, Morton SC, Shekelle PG: **Systematic reviews for evidence-based management: how to find them and what to do with them.** *Am J Manag Care* 2004, **10**:806-12.
- 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.
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
- Egger M, Juni P, Bartlett C, et al.: **How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.** *Health Technology Assessment* 2003, **7**(1): [<http://www.ncchta.org/execsumm/summ701.htm>].

37. Gartlehner G, West SL, Lohr KN, Kahwati L, Johnson JG, Harris RP, et al.: **Assessing the need to update prevention guidelines: a comparison of two methods.** *Int J Qual Health Care* 2004, **16**:399-406.
38. Royle P, Milne R: **Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches.** *Int J Technol Assess Health Care* 2003, **19**:591-603.
39. Royle P, Waugh N: **Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.** *Health Technology Assessment* 2003, **7**(34): [<http://www.ncchta.org/execsumm/summ734.htm>].
40. Beahler CC, Sundheim JJ, Trapp NI: **Information retrieval in systematic reviews: challenges in the public health arena.** *Am J Prev Med* 2000, **18**(4 Suppl):6-10.
41. Ogilvie D, Hamilton V, Egan M, Petticrew M: **Systematic reviews of health effects of social interventions: I. Finding the evidence: how far should you go?** *J Epidemiol Community Health* 2005, **59**:804-8.
42. Alpi KM: **Expert searching in public health.** *J Med Libr Assoc* 2005, **93**:97-103.
43. French SD, McDonald S, McKenzie JE, Green SE: **Investing in updating: how do conclusions change when Cochrane systematic reviews are updated?** *BMC Medical Research Methodology* 2005, **5**:33.
44. Bastian H, Doust J: **When does an updated meta-analysis have enough content to justify re-reading?** [abstract]. Barceona: XI Cochrane Colloquium. October 26-31, 2003.
45. Johnston ME, Brouwers MC, Browman GP: **Keeping cancer guidelines current: results of a comprehensive prospective literature monitoring strategy for twenty clinical practice guidelines.** *Int J Technol Assess Health Care* 2003, **19**:644-55.
46. Barrowman NJ, Fang M, Sampson M, Moher D: **Identifying null meta-analyses that are ripe for updating.** *BMC Medical research Methodology* 2003, **3**:1.
47. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, Woolf SH: **Validity of the Agency for Healthcare Research and Quality Clinical Practice Guidelines. How quickly do guidelines become outdated?** *JAMA* 2001, **286**:1461-7.
48. Stevens A, Colin-Jones D, Gabbay J: **'Quick and clean': authoritative health technology assessment for local health care contracting.** *Health Trends* 1995, **27**:37-42.
49. Hailey D, COrabian P, Harstall C, Schneider W: **The use and impact of rapid health technology assessments.** *Int J Technol Assess Health Care* 2000, **16**:651-6.
50. Perleth M, Luhmann D, Givis B, Droste S: **Rapid Assessments – quick evaluation of medical technology.** *Gesundheitswesen* 2001, **63**(Suppl 1):S79-84.
51. Canadian Coordinating Office for Health Technology Assessment: **Rapid response services in health technology assessment: information from INAHTA members and additional information.** Ottawa: CCOHTA; 2004.
52. Milne R, Clegg A, Stevens A: **HTA responses and the classic HTA report.** *J Public Health Med* 2003, **25**:102-6.
53. Piehl JH, Green S, Silagy C: **Training practitioners in preparing systematic reviews: a cross-sectional survey of participants in the Australasian Cochrane Centre training program.** *BMC Health Services Research* 2002, **2**:11.
54. GRADE Working Group: **Grading quality of evidence and strength of recommendations.** *BMJ* 2004, **328**:1490.
55. Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 9. Grading evidence and recommendations.** *Health Res Policy Syst* 2006.
56. Vist G: **Summaries of findings in Cochrane reviews.** *The Cochrane Collaboration Methods Groups Newsletter* 2005, **9**:5-6 [http://www.cochrane.org/newslett/MGNews_June_2005.pdf].
57. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF: **Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses.** *Lancet* 1999, **354**:1896-900.
58. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
59. Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 10. Integrating values and consumer involvement.** *Health Res Policy Syst* 2006.
60. Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 5. Group processes.** *Health Res Policy Syst* 2006.
61. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, et al.: **Consensus development methods, and their use in clinical guideline development.** *Health Technol Assess* 1988, **2**.
62. Edejer TTT: **Improving the Use of Research Evidence in Guideline Development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications.** *Health Res Policy Syst* 2006.
63. Guyatt G, Yayward R, Richardson WS, Green L, Wilson M, Sinclair J, et al.: **Moving from evidence to action.** In *Users' Guides to the Medical Literature* Edited by: Guyatt G, Rennie D. Chicago: AMA Press; 2002:175-99.
64. **Agency for Healthcare Research and Quality. Updating Systematic Reviews. Expected Release Date: early 2007** [<http://www.ahrq.gov/clinic/tp/sysrevtp.htm>]
65. Moher D: **When and How to Update Systematic Reviews: A Methodological Systematic Review and International Conference.** Canadian Coordinating Office for Health Technology Assessment. 2005 HTA Capacity Building Grants Program. [https://www.ccohta.ca/HTA/hta_grants_2005_descriptions_e.cfm].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations

Holger J Schünemann*¹, Atle Fretheim² and Andrew D Oxman²

Address: ¹INFORMA/CLARITY Research Group, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and
²Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Holger J Schünemann* - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@kunnskapssenteret.no; Andrew D Oxman - oxman@online.no
* Corresponding author

Published: 05 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:21 doi:10.1186/1478-4505-4-21

Accepted: 05 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/21>

© 2006 Schünemann et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the ninth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on grading evidence and recommendations in guidelines.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct a full systematic review ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: Should WHO grade the quality of evidence and the strength of recommendations?

- Users of recommendations need to know how much confidence they can place in the underlying evidence and the recommendations. The degree of confidence depends on a number of factors and requires complex judgments. These judgments should be made explicitly in WHO recommendations. A systematic and explicit approach to making judgments about the quality of evidence and the strength of recommendations can help to prevent errors, facilitate critical appraisal of these judgments, and can help to improve communication of this information.

What criteria should be used to grade evidence and recommendations?

- Both the quality of evidence and the strength of recommendations should be graded. The criteria used to grade the strength of recommendations should include the quality of the underlying evidence, but should not be limited to that.
- The approach to grading should be one that has wide international support and is suitable for a wide range of different types of recommendations. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which is currently suggested in the Guidelines for WHO Guidelines, is being used by an increasing number of other organizations internationally. It should be used more consistently by WHO. Further developments of this approach should ensure its wide applicability.

Should WHO use the same grading system for all of its recommendations?

- Although there are arguments for and against using the same grading system across a wide range of different types of recommendations, WHO should use a uniform grading system to prevent confusion for developers and users of recommendations.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the ninth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

For over 25 years a growing number of organisations have employed various systems to grade the quality of evidence (sometimes called levels of evidence) and the strength of recommendations [1]. Unfortunately, different organisations use various grading systems, which may lead to confusion among consumers.

Groups making recommendations always make judgements about the quality of evidence and the balance of benefits and downsides (harms, burden and costs). Frequently these judgements are made implicitly rather than explicitly and judgements about the quality of evidence are confused with judgements about the balance of benefits and downsides. Many systems that are used to grade the quality of evidence and the strength of recommendations also confuse these judgements by equating the strength of recommendation with the quality of evidence, for example by grading recommendations for which there is high quality evidence as strong, without explicitly considering the balance of benefits and downsides.

Knowing the quality of evidence is essential, but not sufficient for making judgements about the strength of a recommendation. For instance, high quality evidence from well executed randomized controlled trials showed that oral anticoagulation administered for more than one year reduces the risk for recurrent thromboembolic events in patients after a first episode of spontaneous deep venous thrombosis. However, because oral anticoagulation is associated with harms (bleeding risk), burden (taking medication and monitoring anticoagulation levels) and cost (anticoagulation clinics or monitoring devices) the recommendation to anticoagulate all patients is weak because the benefits and downsides are finely balanced and individual patients will make different choices [2]. Both judgements about the quality of evidence and about the strength of a recommendation are complex and require consideration of a number of factors.

In this paper we addressed the following questions:

- Should WHO grade the quality of evidence and the strength of recommendations?

- What criteria should be used to grade evidence and recommendations?

- Should WHO use the same grading system for all of its recommendations?

Questions related to what evidence should be included, how it should be synthesized and reported are addressed in other papers in this series [3-5].

What WHO is doing now?

WHO groups (e.g. WHO Europe) have acknowledged the need for evaluating or developing a grading system [6,7], and the Guidelines for WHO Guidelines recommend using a specific, uniform grading system [8]. However, this system, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, has scarcely been used within WHO [9,10]. Some WHO groups have developed their own grading systems [11,12], despite of the guidelines for WHO guidelines suggestion to use GRADE. Most have not explicitly graded either the quality of evidence or the strength of recommendations [13,14].

What other organisations are doing

Most, but not all organizations that develop guidelines use a grading system to express the strength of a recommendation or the quality of evidence. For example, the US Preventive Services Task Force (USPSTF) uses a grading system that assigns one of three grades of evidence: good, fair, or poor [15]. The Task Force uses its assessment of the evidence and magnitude of net benefit to make a recommendation, coded as a letter: from A (strongly recommended) to D (recommend against). The UK National Institute for Health and Clinical Excellence (NICE) has not yet made a decision as to which grading system to use [16]. The Scottish Intercollegiate Guideline Network (SIGN) has developed its own grading system for application to SIGN guidelines [17]. The Australian Medical Research Council is currently developing a grading system that will probably include grading recommendations according to strength of recommendations and quality of evidence [18]. The US Task Force on Community Preventive Services uses a system in which the quality of the evidence of effectiveness links directly the strength of the recommendation [19,20]. Professional organizations use a variety of systems, many of them, however, based on two prominent grading approaches: the system derived from the Canadian Task Force on the Periodic Health Examination [1,21] and a successor of that system, the Oxford Centre for Evidence Based Medicine approach [22].

More recently, medical societies have begun to form collaborations within specialties to develop grading systems

on their own. For example a group of specialty societies in rehabilitation sciences formed a panel to develop an approach to grading the quality of evidence and strength of recommendations [23]. This panel developed a set of criteria for grading the strength of both the evidence and the recommendation. Similarly, the world leading urology associations have come together to adopt a uniform grading system and approach that would be useful for urologists around the world rather than each association using a different grading system [24]. This latter collaboration named, Evidence Based Urology, is exploring using the GRADE approach. The GRADE approach is being used increasingly by organisations around the world [25-28], although in some cases with slight modifications [29]. It has been used for public health questions such as the pharmacological management of human influenza A(H5N1) infection (avian flu) [30], although it more commonly has been used for clinical questions up to now. A group of family practice and primary care journals has also developed a system to grade the strength of a recommendation [31].

Methods

The methods used to prepare this review are described in the introduction to this series [32]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We reviewed existing guidelines for guidelines to identify grading system currently in use. We also searched PubMed using (grading system) and (methods) (MESH headings/keywords) for systematic reviews and studies of methods for grading the quality of evidence. In addition, we searched databases maintained by the Agency for Healthcare Research and Quality (AHRQ, [33]) and the Guidelines International Network (GIN, [34]). These searches were supplemented with information obtained directly from guideline development organizations and our own files. Because of our involvement with organizations that produce guidelines and prior work with grading systems, in particular the GRADE system, we had in depth knowledge about several systems [25,28,29,35,36].

Findings

We identified one systematic review dealing with the evaluation of grading systems. In 2002, the US Agency for Healthcare Research and Quality (AHRQ) published a systematic review of existing systems to grade the quality of evidence and strength of recommendations [37]. The AHRQ review considered 40 systems until the year 2000 that addressed grading the strength of a body of evidence. The important domains and elements for the systems to grade the strength of evidence that the authors agreed on were quality (the aggregate of quality ratings for individ-

ual studies, predicated on the extent to which bias was minimized), quantity (magnitude of effect, numbers of studies, and sample size or power) and consistency (for any given topic, the extent to which similar findings are reported using similar and different study designs).

More recently, independent work by the Canadian Optimal Medication Prescribing and Utilization Service (COMPUS) used a detailed process to evaluate and select an evidence grading system and expanded the work by AHRQ (while accepting it) until the year 2005 [38]. COMPUS, which identifies, evaluates, promotes and facilitates best practices in drug prescribing and use among health care providers and consumers in Canada [39], is a nationally coordinated program, funded by Health Canada and delivered by the Canadian Coordinating Office for Health Technology Assessment (CCOHTA, [39]). They assembled a working group of internal researchers, information specialists, methodology experts, and external researchers to update the work of AHRQ. COMPUS searched for and selected review articles for the period 2000 to 2005. This resulted in more than 3,000 citations for selection. Eleven review articles were selected for further analysis based on a priori selection criteria specified by the working group. Nearly 50 evidence grading systems were identified from the 11 review articles. Canadian and international experts in evidence evaluation methodology helped identify an additional 10 instruments or systems not included in the list of identified grading systems. The identified instruments and systems were evaluated using the AHRQ evaluation grids. The highest scoring instruments were the GRADE and the SIGN approach [38]. A second round of expert consultation and stakeholder input from all interested parties confirmed the selection of these instruments.

The GRADE system was developed through an intensive international collaboration of methodologists, guideline developers and clinicians and incorporates the factors identified in the AHRQ review and described above [35,36].

Should WHO grade the quality of evidence and the strength of recommendations?

We did not identify published studies that compared graded with non-graded recommendations. The only evidence we are aware of are three unpublished studies. The first was conducted by UpToDate[®], an electronic textbook, that asked a small group of users to compare graded with non-graded recommendations and explore – in a focus group setting – reasons for their answers (UpToDate[®], personal communication). The second is our own study asking a small group of the general public interested in health care issues (Akl E, et al, manuscript in preparation). The third is a study by researchers in Norway who provided patients with back problems with graded evidence of the

effects of alternative interventions graded with the GRADE approach. Users of the website intuitively understood the meaning of the quality grades for each outcome (Claire Glenton, personal communication). The findings of these evaluations suggested that users preferred graded over non-graded recommendations.

Despite the lack of stronger direct evidence, there is agreement among most guideline developers that grading the quality of evidence has advantages, because health care decisions involve a trade-off between likely benefits on the one hand, and downsides (harms, burden and costs) on the other hand [40]. To integrate these recommendations with their own judgment, guideline users need to understand the basis for the recommendations that guidelines offer them. A systematic approach to grading the strength of recommendations should minimize bias and aid interpretation about benefits and downsides. In addition, a systematic and explicit approach to making judgments about the quality of evidence and the strength of recommendations is likely to help prevent errors, facilitate critical appraisal of these judgements, and can help improve communication of this information [36].

What criteria should be used to grade evidence and recommendations?

In a series of 16 international meetings and correspondence over five years the GRADE Working Group has derived a set of criteria to assess the quality of evidence (Table 1) and the strength of recommendations (Table 2) [25,29,35,36,41,42]. The GRADE system has several advantages over other systems including explicit definitions and sequential judgments during the grading process; a detailed description of the criteria for the quality of evidence for single outcomes and for the overall quality of the evidence; weighing the relative importance of outcomes; consideration of the balance between health benefits versus harms, burdens and cost; and the development of evidence profiles and summaries of find-

ings. In addition the GRADE group is supported by an international collaboration [36]. The main limitation and criticism of the GRADE system is its complexity. Work in progress is addressing this limitation including the development of user friendly software to develop evidence profiles (G. Vist, personal communication and [26,29]).

Should WHO use the same grading system for all of its recommendations?

We did not identify evidence for or against using a single grading system for all types of recommendations, including clinical, public health and health policy recommendations. The arguments for and against using a single grading system are summarised in Table 3. The most important reasons for using a consistent system are a) minimising confusion amongst users of WHO recommendations; b) the risk of bias if groups can select a system that makes the quality of evidence and the strength of recommendations look better for their preferred interventions; and c) being intellectually honest about recognising the limits of the evidence rather than having a double standard. If an approach can be identified that is suitable across a wide range of interventions and contexts both methodologically and politically, the advantages outweigh the disadvantages.

Some developers and users of GRADE believe that GRADE can be consistently and usefully applied across clinical and non-clinical interventions, based on conceptual arguments and experience up to now applying this approach to a wide range of interventions, including public health and health system interventions. Others disagree because they believe it is unlikely to be an appropriate approach for some areas for the reasons summarised in table 3. There is not yet an empirical evidence base with which to mediate this disagreement for GRADE or any other grading system. Up to now GRADE has been used mostly for clinical interventions and few examples of its use with public health questions have been published. There is an

Table 1: GRADE quality assessment criteria

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

* 1 = move up or down one grade (for example from high to intermediate) 2 = move up or down two grades (for example from high to low)
 ** A statistically significant relative risk of >2 (< 0.5), based on consistent evidence from two or more observational studies, with no plausible confounders
 *** A statistically significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity

Table 2: Decisions about the strength of a recommendation

Factors that can weaken the strength of a recommendation	Explanation
Lower quality evidence	Will create greater uncertainty about the size of the (relative) effects (benefits and harms)
Uncertainty about the balance of benefits versus harms and burdens	Uncertainty about the baseline risk, prevalence of a problem or health status, which could affect the size of the (absolute) effects
Uncertainty or differences in values	Uncertainty about the relative importance of the benefits and downsides to those affected, or differences in how important they are to different people, which could affect the balance between the benefits versus harms and burden
Marginal net benefits or downsides	The anticipated net benefits or downsides are small (and uncertain)
Uncertainty about whether the net benefits are worth the costs	Uncertainty related to lack of information about the cost or whether the resource expenditure is justified by the anticipated benefit

ongoing international collaborative effort to apply the GRADE approach to public health and health systems interventions, and it is possible that modifications may be needed to ensure its usefulness for non-clinical interventions. For example, in one recent review of drug policies the authors felt that it was important to distinguish between different types of observational studies (interrupted time-series analyses and controlled before-after studies) when making judgements about the quality of evidence for important outcomes [43].

Discussion

WHO has made a decision to use a grading system to grade the quality of evidence and strength of recommendations that is sensible and is being used widely, the GRADE system [36]. WHO has been involved in the development of this system from the beginning, and consideration has been given to the potential for application of the system to WHO guidelines in developing the GRADE approach. This might have been expected to facilitate the dissemination and adoption of this approach by WHO guideline developers. However, interest in GRADE workshops at WHO has been limited, there is not a tradition of grading the quality of evidence or strength of recommendations at WHO, and few resources have been invested in supporting the use of GRADE specifically, or in supporting more rigorous guidelines development methods generally.

More recently, however, the WHO rapid advice guideline panel for the pharmacological management of human infection with avian influenza A (H5N1) virus applied GRADE successfully [30] and several WHO guidelines are under development using GRADE (Sue Hill, personal communication). In general, the evidence that graded recommendations have advantages over non-graded recommendations is limited, but there are strong arguments, including the clear and transparent communication of how much confidence users can place in recommendations and the evidence underlying them. Another limitation is that both the quality of evidence and the strength of recommendations exist on a continuum. Categorization of quality into four categories and recommendations for or against treatments into two grades, strong and weak, may oversimplify complex health care recommendations, but guidelines consumers are generally likely to benefit from this simplification as they are most interested in which recommendations to follow.

Further work

We have found a large body of work on the development and evaluation of various grading systems. Problems have arisen because many different grading systems exist. Future efforts should focus on forging a consensus on using a sensible and uniform approach to grade the quality of evidence and strength of recommendations, building on the work of the GRADE working group. Use of the

Table 3: Pros and cons of using the same system for grading evidence and formulating recommendations for a wide range of health care interventions, including clinical and non-clinical interventions

Arguments for having a common approach	Arguments against having a common approach
<ul style="list-style-type: none"> • Having less demanding systems for some kinds of questions might result in false positive conclusions. • People with vested interests in particular interventions could choose the system that makes their intervention look best. • People with vested interests in particular evaluation approaches could choose the system that makes their preferred evaluation approach look best. • Having different systems for different types of interventions might be confusing. • It is intellectually honest to recognise the limits of evidence where this is appropriate. • Admitting the limitations of evidence, if this is appropriate, might promote more and better research. 	<ul style="list-style-type: none"> • Having an infeasible system for some kinds of questions might result in false negative conclusions. • False negative conclusions due to inappropriate evaluation requirements may have negative political and health consequences; for example, effective programs that cannot be studied with randomised trials might experience funding cuts. • Interventions that cannot be studied with randomised trials might not be evaluated. • A single system may not discriminate adequately within the range of evidence that is appropriate to consider for clinical and non-clinical interventions. • A system that can adequately address evidence across a wide range of interventions and contexts may be overly complex.

GRADE system by WHO, as is currently recommended by the Guidelines for WHO Guidelines, could help by obtaining more experience, particularly with non-clinical interventions, contribute to improvements in the existing system, contribute to agreeing on a common international approach to grading of recommendations and help to ensure the quality and transparency of the judgements that are made across various groups that make recommendations on behalf of WHO. Development of software and a detailed manual to simplify the use of the GRADE system is underway and should facilitate the use of this system and its further development.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

HJS prepared the first draft of this review. AF and ADO authors contributed to drafting and revising it. All authors read and approved the final manuscript

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper. The authors thank Mr. John Fox for administrative assistance with searching and maintaining databases.

References

1. Anonymous: **Canadian Task Force on the Periodic Health Examination. The periodic health examination.** *CMAJ* 1979, **121**:1193-1254.
2. Buller HR, Agnelli G, Hull RD, Hyers TM, Prins MH, Raskob GE: **Antithrombotic therapy for venous thromboembolic disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy.** *Chest* 2004, **126**(3 Suppl):401S-428S.
3. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
4. Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006.
5. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 14. Reporting guidelines.** *Health Res Policy Syst* 2006.
6. **The Russian Federation. Making Pregnancy Safer/Promoting Effective Perinatal Care. National Orientation and Planning Meeting. Moscow, 11-14 February 2003.**
7. **Fourth Futures Forum of High-Level Decision-Makers Tools for decision-making in public health. Brussels, Belgium 16-17 June 2003.**

8. **Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines. EIP/GPE/EQC/2003.1.** Geneva, World Health Organization; 2003.
9. Schünemann HJ, Hill S, Kakad M, Bellamy R, Uyeki T, Hayden F, Yazdanpanah Y, Beigel J, Chotpitayusunondh T, Vist G, Del Mar C, Farrar J, Fukuda K, Hien TT, Shindo N, Stockman L, Sugaya N, Tomson G, Nagjdaliyev A, Özbay B, Roth C, Croisier A, Oxman, AD: **WHO Rapid Advice Guidelines for the pharmacological management of human infection with avian influenza A (H5N1) virus.** *Lancet companion paper*: submitted
10. Schünemann HJ, Hill S, Kakad M, Vist G, Bellamy R, Stockman L, Uyeki T, Del Mar C, Shindo N, Fukuda K, Hayden F, Yazdanpanah Y, Oxman, AD: **WHO Rapid Advice Guidelines: Quick and Transparent.** *Lancet companion paper*: submitted
11. **World Health Organization. Prevention of Recurrent Heart Attacks and Strokes in Low and Middle Income Populations: Evidence-based Recommendations for Policy Makers and Health Professionals.** Geneva, World Health Organization; 2003.
12. **Guidelines for the Prevention of Malaria.** Geneva, World Health Organization; 2006.
13. Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.** (submitted)
14. Panisset U: **A Review of WHO recommendations published in 2005.** *Health Res Policy Syst* 2006.
15. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D: **Current methods of the U.S. Preventive Services Task Force: A review of the process.** *American Journal of Preventive Medicine* 2001, **20**(3, Supplement 1):21-35.
16. **National Institutes for Health and Clinical Excellence** [http://www.nice.org.uk/pdf/GDM_Allchapters_0305.pdf]
17. Harbour R, Miller J: **A new system for grading recommendations in evidence based guidelines.** *BMJ* 2001, **323**:334-336.
18. Middleton P, Salisbury J, Coleman K, Norris S, Grimmer K, Hillier S: **Assessing the body of evidence and grading recommendations in evidence-based clinical practice guidelines: Melbourne, Australia.** Edited by: Collaboration C.; 2005:42.
19. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Agüero L, Truman BI, Hopkins DP, Mullen PD, Thompson RS, Woolf SH, Carande-Kulis VG, Anderson L, Hinman AR, McQueen DV, Teutsch SM, Harris JR: **Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services.** *Am J Prev Med* 2000, **18**(1 Suppl):35-43.
20. Truman BI, Smith-Akin CK, Hinman AR, Gebbie KM, Brownson R, Novick LF, Lawrence RS, Pappaioanou M, Fielding J, Evans CA Jr., Guerra FA, Vogel-Taylor M, Mahan CS, Fullilove M, Zaza S: **Developing the Guide to Community Preventive Services--overview and rationale. The Task Force on Community Preventive Services.** *Am J Prev Med* 2000, **18**(1 Suppl):18-26.
21. Sackett DL: **Rules of evidence and clinical recommendations on the use of antithrombotic agents.** *Archives Int Med* 1986, **146**:464-465.
22. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M.: **Levels of Evidence and Grades of Recommendation. Centre for Evidence-Based Medicine.**
23. **Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions: overview and methodology.** *Phys Ther* 2001, **81**(10):1629-1640.
24. **Evidence Based Urology Task Force** [<http://www.gradeworkinggroup.org/EBUro/intro.htm>]
25. Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schünemann H: **An emerging consensus on grading recommendations? (Editorial).** *ACP J Club* 2006, **144**(1):A08.
26. **GRADE Working Group** [<http://www.gradeworkinggroup.org>]
27. **Researchers test system to custom-tailor guidelines.** *Clin Resour Manag* 2002, **3**(1):8-10, 1.
28. Schünemann HJ, Jaeschke J, Cook D, Bria WF, El Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer J, McNicholas W, Rubenfeld G, Turino GM, Guyatt G (on behalf of the ATS Documents Development and Implementation Committee): **An official ATS statement: Grading the quality of evidence and strength of recommendations for guidelines and recommendations.** *American Journal of Respiratory and Critical Care Medicine* 2006, **174**(5):605-614.
29. Guyatt G, Gutterman D, Baumann M, Addrizzo-Harris D, Hylek E, Phillips B, Raskob G, Lewis S, Schünemann H.: **Grading strength of**

- recommendations and quality of evidence in clinical guidelines. *Chest* 2006, **129**:174-181.
30. **WHO Rapid Advice Guidelines on pharmacological management of humans infected with avian influenza A (H5N1) virus** [http://www.who.int/csr/disease/avian_influenza/guidelines/en/index.html]
 31. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, Bowman M: **Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature.** *J Am Board Fam Pract* 2004, **17(1)**:59-67.
 32. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006.
 33. **Agency for Healthcare Research and Quality** [<http://www.guidelines.gov>]
 34. **Guidelines International Network** [<http://www.g-i-n.net>]
 35. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr.: **Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group.** *BMC Health Serv Res* 2004, **4(1)**:38.
 36. 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, Schunemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr., Zaza S: **Grading quality of evidence and strength of recommendations.** *Bmj* 2004, **328(7454)**:1490.
 37. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L: **Systems to Rate the Strength of Scientific Evidence.** In *Evidence Report/Technology Assessment Volume AHRQ Publication No. 02-E016.* Rockville, MD, Agency for Healthcare Research and Quality; 2002.
 38. **Canadian Optimal Medication Prescribing and Utilization Service** [https://www.ccohta.ca/compus/compus_pdfs/COMPUS_Evaluation_Methodology_draft_e.pdf]
 39. **Canadian Coordinating Office for Health Technology Assessment** [<https://www.ccohta.ca/>]
 40. Schünemann HJ, Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 6. Determining which outcomes are important.** *Health Res Policy Syst* 2006.
 41. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr.: **Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system.** *BMC Health Serv Res* 2005, **5(1)**:25.
 42. Schünemann HJ, Best D, Vist G, Oxman AD: **Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations.** *Cmaj* 2003, **169(7)**:677-680.
 43. Aaserud M, Dahlgren A, Kusters JP, Oxman AD, Ramsay C, Sturm H.: **Pharmaceutical policies: effects of pricing and purchasing policies on rational drug use.** In *The Cochrane Library* Chichester, UK, John Wiley & Sons, Ltd; 2006.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement

Holger J Schünemann*¹, Atle Fretheim² and Andrew D Oxman²

Address: ¹INFORMA/CLARITY Research Group, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and
²Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Holger J Schünemann* - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@kunnskapssenteret.no ; Andrew D Oxman - oxman@online.no
* Corresponding author

Published: 05 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:22 doi:10.1186/1478-4505-4-22

Accepted: 05 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/22>

© 2006 Schünemann et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 10th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on integrating values and consumers in guideline development.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We reviewed the titles of all citations and retrieved abstracts and full text articles if the citations appeared relevant to the topic. We checked the reference lists of articles relevant to the questions and used snowballing as a technique to obtain additional information. We did not conduct a full systematic review ourselves. Our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We did not find a systematic review of methods for integrating values in guidelines, but we found several systematic reviews that dealt with related topics.

Whose values should WHO use when making recommendations?

- Values, the relative importance or worth of a state or consequences of a decision (outcomes relating to benefits, harms, burden and costs), play a role in every recommendation. Ethical considerations, concepts that determine what is right, also play a role.
- The values used in making recommendations should reflect those of the people affected. Judgements should be explicit and should be informed by input from those affected (including citizens, patients, clinicians and policy makers).
- When differences in values may lead to different decisions or there is uncertainty about values, this should also be explicit. If differences in values are likely to affect a decision, such that people in different setting would likely make different choices about interventions or actions based on differences in their values, global recommendations should be explicit in terms of which values were applied and allow for adaptation after incorporating local values.

How should WHO ensure that appropriate values are integrated in recommendations?

- All WHO guideline groups should uniformly apply explicit, transparent and clearly described methods for integrating values.
- WHO should consider involving relevant stakeholders if this is feasible and efficient.

- WHO should develop a checklist for guidelines panels to help them to ensure that ethical considerations relevant to recommendations are addressed explicitly and transparently.

How should users and consumers be involved in generating recommendations?

- Including consumers in groups that are making global recommendations presents major challenges with respect to the impossibility of including a representative spectrum of consumers from a variety of cultures and settings. Nonetheless, consideration should be given to including consumers in groups who are able to challenge assumptions that are made about the values used for making recommendations, rather than represent the values of consumers around the world.
- WHO should establish a network to facilitate involvement of users.
- Draft recommendations should be reviewed by consumers, who should be asked explicitly to consider the values that were used.

How should values be presented in recommendations?

- Recommendations should include a description of how decisions were made about the relative importance of the consequences (benefits, harms and costs) of a decision.
 - Values that influence recommendations should be reported along with the research evidence underlying recommendations.
 - When differences in values would lead to different decisions or there is important uncertainty about values that are critical to a decision, this should be flagged and reflected in the strength of the recommendation.
 - Adaptable guideline templates that allow for integration of different values should be developed and used when differences in values are likely to be critical to a decision.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 10th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Utilities, health state preferences or values are the desirability or preference that individuals exhibit for a particular health state [1]. Individuals usually assign less value to and have less preference for more impaired health states (e.g. death or dependency after a stroke) compared to other health states (e.g. full health or having a very mild stroke without serious sequela). In this document we will use the terms "values" to refer to the relative worth or importance of a health state or consequences (benefits, harms and costs) of a decision. It is primarily this concept of values that we focus on here.

Ethical or moral values also play a role in making health care recommendations [2-4]. These refer to concepts of what is right based on philosophical, humanistic or religious considerations. Ethical values can vary among individuals within a society and across societies or culture, and may influence recommendations and the implemen-

tation of recommendations. We will refer to these as ethical considerations.

Several formal methods exist to measure values in health-care [5,6]. The principle methods are based on direct utility instruments including the standard gamble, time-trade off and visual analogue scales [5,6]. Direct preference-based instruments generate a value score for respondents' current health state or hypothetical states, typically on a 0.0 to 1.0 scale where 0.0 indicates dead and 1.0 indicates full health. Indirect methods offer alternatives to direct assessments and include multi-attribute utility tools and transformations based on quality of life assessments [7-10]. Multi-attribute utility tools ask respondents to describe their health state, and the value of that health state is calculated using a developed formula representing preferences of the general population. The Euroquo, Quality of Well-Being Index, and the Health Utilities Index are examples of this approach [9-11]. Transformations from generic health related quality of life tools utilize modelling techniques that transform quality of life scores into values [8]. Ranking scales and qualitative methods compliment these methods [6]. However, the application of their results is complicated by the fact that the reproducibility between the different methods is poor and each of the methods has strength and limitations.

Values play a role in every recommendation, either explicitly or implicitly. For instance patients who suffered an

idiopathic deep venous thrombosis (DVT) usually receive treatment with adjusted dose warfarin for one year to prevent recurrent DVT and pulmonary embolism [12]. Continuing on standard-intensity warfarin beyond the treatment of one year will reduce his absolute risk for recurrent DVT by more than 7% per year for several years [13]. The burdens of treatment include taking a warfarin pill daily, keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. Patients who are very averse to a recurrent DVT would consider the benefits of avoiding DVT worth the downsides of taking warfarin. Other patients are likely to consider the benefit not worth the harms and burden. Another example refers to the different values patients with atrial fibrillation and their clinicians place on the adverse consequences of stroke and gastrointestinal bleed [14]. A third example relates to a health care recommendation about the combination of chemotherapy and radiotherapy versus radiotherapy alone in unresectable, locally advanced non-small cell lung cancer [15,16]. Compared with radiotherapy alone, the combination of chemotherapy and radiotherapy reduces the risk for death corresponding to a mean gain in life expectancy of a few months [15], but increases harm and burden related to chemotherapy. Thus, considering the values and preferences patients would place on the small survival benefit in view of the harms and burdens, guideline panels may offer a weak recommendation despite the high quality of the available evidence. Generally, there is agreement that the values that are used for comparing the relative benefits and downsides of interventions should be explicit [17-20].

In this paper we addressed the following questions:

- Whose values should WHO use when making recommendations?
- How should WHO ensure that appropriate values are integrated in recommendations?
- How should users and consumers be involved in generating recommendations?
- How should values be presented in recommendations?

Questions related to identifying important outcomes, group composition and equity considerations are addressed in other papers in this series [21-23].

What WHO is doing now

The Guidelines for WHO Guidelines suggests that end users, and patients specifically, should be represented on guideline panels [24]. However, review of selected WHO

guidelines did not yield information about the inclusion of end users or patients in guideline groups or the use of other methods to ensure appropriate integration of values and consumer involvement. The fact that WHO makes global recommendations presents challenges for obtaining appropriate representation of values, because values may differ across different cultures and settings.

What other organisations are doing?

A number of guideline developers have invited individuals who could represent and understand the perspectives of stakeholders, including consumers [25]. However, in a recent survey of organizations and specialty societies that develop guidelines only approximately 25% regularly involved consumers in the process [26]. In another survey of 18 prominent organizations that develop guidelines approximately 50% regularly involved patients [27].

The UK National Institute for Health and Clinical Excellence (NICE) has adopted a very comprehensive approach to involving patients and consumers and has formed a patient involvement unit aiming to involve patients and carers in the development of individual clinical guidelines [20,28]. NICE consumer involvement can be categorized into four broad areas:

1) Stakeholder consultation

Organisations can register and comment at any stage during the development process. They can nominate patient groups and participate and comment on the development of a guideline at every stage from the suggestion of guideline topics, drafting of scopes, development and initial drafting of guidelines, to the second consultation draft.

2) Direct input

NICE committees and working groups are expected to include at least two members who play a crucial role by providing a patient/carer perspective to their discussions and decisions. They may be patients, carers or patient advocates. Vacancies are publicised via national patient and carer organisations, on a website or via the national press.

3) Indirect input

Examples include focus groups with patients, patient written testimonials and video-taped interviews with patients that were presented to a technology appraisal committee.

4) Dissemination of NICE guidance to and by patients

All NICE guidance is produced in versions written for patients, carers and the public. Copies can be downloaded

from the NICE website or printed copies can be obtained by telephoning the NHS Response line. Patient organisations play an increasingly significant role in helping NICE disseminate its guidance to individual patients and carers and providing feedback.

The Scottish Intercollegiate Guidelines Network (SIGN) involves patients and patient representatives in the guideline development process. For each guideline, SIGN aims to search for evidence about patient important outcomes after having identified a guideline topic, but this is not done consistently (R. Harbour, personal communication). SIGN also aims to include patient representatives and guideline users in the process at all stages during the development process [17]. The UK National Health System (NHS) Health Technology Assessment program also systematically involves public advocates in their work [29-32]. Various speciality societies describe that representation of consumers on guideline panels (at least one member) is required which is often described as representation by a patient advocacy group [33].

The Cochrane Collaboration, which produces systematic reviews, but does not make recommendations, has made consumer involvement an integral part of its work of producing and disseminating systematic reviews and demonstrated the feasibility of international consumer involvement projects. In 1998 approximately two thirds of Cochrane Collaboration review groups had consumer involvement [34]. There is also a Cochrane Consumers Network [35], which helps to provide consumer input into developing Cochrane reviews.

Methods

The methods used to prepare this review are described in the introduction to this series [36]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We reviewed existing guidelines for guidelines to identify processes for integrating consumer values and consumer involvement. We based the current summary and recommendation on the work of prominent developers of guidelines. We also searched PubMed using "consumer" and "involvement" as search terms (MESH headings/keywords) for systematic reviews. We searched PubMed for systematic reviews on how guideline groups integrate values and preferences using the terms "guideline" and "values OR preferences OR utilities" (we identified 694 citations labelled as systematic reviews). We also searched the Cochrane library, Methodology registry and database using the keywords "guideline" and ("values" or "preferences"). We searched databases maintained by the Agency for Healthcare Research and Quality (AHRQ, [37]) and the Guidelines International Network (GIN, [38]), and reviewed infor-

mation obtained from various organizations and our own files. We did not conduct a full systematic review. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Findings

We did not find a systematic review of methods for integrating values in guidelines, but we found one systematic review that compared whether values differ between the general population and patients [39]. Another systematic review that focused on interventions to promote consumer involvement in developing healthcare policy and clinical practice guidelines did not find any comparative studies of consumer involvement versus no consumer involvement or of different ways of involving consumers [40]. We identified systematic reviews dealing with indirect evidence focusing on consumer involvement, for example in research agenda setting [31]. One review focused on involving patients in the planning and development of health care and another systematic review addressed which methods should be used to include the views of the public in policy documents [6,41]. We also found a review summarized values obtained by rating scale, time-trade off and standard gamble for different disease states [42]. In addition, we identified several articles that addressed whether patient values should be integrated in clinical practice guidelines [43].

Whose values should WHO use when making recommendations?

Clinicians' values for health states differ from those of patients and among different clinician groups [14,44]. For example, the values physicians assign to stroke as an outcome and to adverse consequences (e.g., gastrointestinal bleeding) of treatment to prevent stroke in patients with atrial fibrillation differ from those of patients [14]. The values used in making recommendations should reflect those of the people affected. While there is widespread belief that values for health states also differ between patients and the general public, this belief is not supported by the available evidence. A systematic review of 33 studies found that preferences for hypothetical health states did not differ between patients and the public [39].

Since guidelines, for the most part, affect the use of limited public resources and, therefore, inevitably affect the general public, WHO guidelines should consider societal values and recognise when there may be important, legitimate differences in values across different cultures and settings. Judgements should be explicit and should be informed by input from those affected (including citizens, patients, clinicians and policy makers). Representation of all potentially relevant societies in groups developing global recommendations is not feasible. If differences in val-

ues are likely to affect a decision, such that people in different settings would likely make different choices about interventions or actions based on differences in their values, global recommendations should be explicit in terms of which values were applied and allow for adaptation after incorporating local values. It may also be possible to include consumers or other stakeholders in panels with the primary responsibility of questioning assumptions that are made about values, rather than representing the values of any particular group.

We found discussions of ethical considerations in guidelines and health technology assessments [2-4], but we did not find a systematic review of processes for addressing ethical considerations systematically and transparently. There is no standard way for doing this, although Hofmann has taken a step towards developing a practical approach by identifying relevant questions that could be asked in the context of health technology assessments [4].

How should WHO ensure that appropriate values are integrated in recommendations?

The evidentiary basis for appropriate identification and representation of values in guideline development is limited. We did not identify a systematic review focusing on different ways of including values in clinical practice guidelines. A review by Ryan and colleagues focused on public preferences for healthcare [6]. The authors concluded that there was no single, best method to gain public opinion. The method must be carefully chosen and rigorously carried out in order to accommodate the question being asked. They recommend conjoint-based methods (including ranking, rating and choice-based), willingness to pay, standard gamble and time trade-off as quantitative techniques and one-to-one interviews, focus groups, Delphi technique and citizens' juries as qualitative techniques. There were only a few studies that conducted direct comparisons of methods at that time (up to 2000) and the review requires updating.

While the evidence about which method should be adopted is inconclusive, all of the methods mentioned above are acceptable. Methods that transform results of generic health related quality of life instruments into value scores can also be used but they have the disadvantage of requiring calculations and are based on mathematical assumptions in developing transformation equations. If data from primary research on societal or patient values are used, guideline panels should make explicit how these values were measured (e.g., they should specify whether a visual analogue scale, standard gamble or other methods were used).

Decision analysis and economic analyses are approaches to explicitly integrating values into guidelines, which may

sometimes be useful [45]. Less formal methods include representation on guideline panels and consultation with consumers to inform judgements about the relative importance of the benefits and downsides of interventions based on rating scales. WHO guideline groups should uniformly apply explicit, transparent and clearly described methods for integrating values.

How should users and consumers be involved in generating recommendations?

Consumers should be part of guideline groups, be able to contribute and be heard. Van Wersch and Eccles evaluated a guideline development program [46]. They described experience with three alternative methods of consumer involvement in guideline development: a) incorporating individual patients; b) a one off meeting with patients; and c) incorporating a consumer advocate in the guideline development group. They concluded that consumers should be involved in all stages of guideline development, while acknowledging that this is not straightforward, that there is no right way to accomplish this, and that more research is required to optimize the process and outcomes.

Indirect evidence about the involvement of consumers comes from studies evaluating the involvement of consumers in research priority setting. Oliver and colleagues systematically reviewed different methods of consumer involvement in research priority setting [31]. They concluded that "what we know about the advantages and disadvantages of methods for involving consumers in agenda setting rests on weak short-term evidence and almost entirely speculative long-term evidence". Telford and colleagues used a Delphi approach to identify principles and indicators of successful involvement of consumers in research [47]. They identified eight principles for the successful involvement in research that could also be used by WHO to ensure that consumers are adequately represented in guideline development projects (Table 1).

Any involvement of consumers requires a clear understanding of the evidence by consumers. Difficulties with medical terminology or other jargon are an important barrier to involvement. Well-informed and experienced consumers are more likely to interact with the guideline developers than those who are less informed or less familiar with medical terminology or other jargon that is used.

A large number of studies have been conducted using a variety of methods to elicit values for direct patient care questions, but this literature is not well summarised and it may be beyond the capacity of most groups developing recommendations to systematically review the literature relevant to the specific questions that they are addressing. The review by Morimoto and Fukui is limited to direct

Table 1: The principles and indicators of successful consumer involvement in NHS research (from Telford et al.) [37]

Principle	Indicator(s)
1 The roles of consumers are agreed between the researchers and consumers involved in the research	• The roles of consumers in the research were documented
2 Researchers budget appropriately for the costs of consumer involvement in research	• Researchers applied for funding to involve consumers in the research • Consumers were reimbursed for their travel costs • Consumers were reimbursed for their indirect costs (e.g. carer costs)
3 Researchers respect the differing skills, knowledge and experience of consumers	• The contribution of consumers-skills, knowledge and experience were included in research reports and papers
4 Consumers are offered training and personal support, to enable them to be involved in research	• Consumers – training needs related to their involvement in the research were agreed between consumers and researchers • Consumers had access to training to facilitate their involvement in the research • Mentors were available to provide personal and technical support to consumers
5 Researchers ensure that they have the necessary skills to involve consumers in the research process	• Researchers ensured that their own training needs were met • in relation to involving consumers in the research
6 Consumers are involved in decisions about how participants are both recruited and kept informed about the progress of the research	• Consumers gave advice to researchers on how to recruit participants to the research • Consumers gave advice to researchers on how to keep participants informed about the progress of the research
7 Consumer involvement is described in research reports	• The involvement of consumers in the research reports and publications was acknowledged • Details were given in the research reports and publications • of how consumers were involved in the research process
8 Research findings are available to consumers, in formats and in language they can easily understand	• Research findings were disseminated to consumers involved in the research in appropriate formats (e.g. large print, translations, audio, Braille) • The distribution of the research findings to relevant • consumer groups was in appropriate formats and easily understandable language • Consumers involved in the research gave their advice on the choice of methods used to distribute the research findings

preference instruments and does not provide values for the full spectrum of diseases for which WHO develops recommendations [42]. However, this and similar reviews can provide some guidance because direct elicitation of values from a representative sample of people is rarely feasible.

WHO should consider developing a database or collaborating with others to establish a database of evidence about the relative values of common health outcomes across different cultures and settings, which could be used to inform the judgements made by groups making recommendations.

How should values be presented in recommendations?

Values should always be considered in making recommendations, although they do not always influence the strength of a recommendation when they are uniformly shared among patients and society. While values should always be made explicit, WHO could restrict its presentation and labelling of values to those that are most important for decision-making. These recommendations could be flagged as being strongly influenced by values and include a presentation of whose values they represent. In particular recommendations should:

- Include a detailed description of how decisions were made about the relative importance of the consequences (benefits and downsides) of a decision. This should routinely be included in the methods section of a guideline.

- Values that influence recommendations should be reported along with the research evidence underlying the recommendations.

- When differences in values would lead to different decisions or there is important uncertainty about values that are critical to a decision, this should be flagged and reflected in the strength of the recommendation.

- Adaptable guideline templates that allow for integration of different values should be developed and used when differences in values are likely to be critical to a decision [48].

Discussion

There is no high quality research informing the choice of whose values guideline panels should use or methods of consumer involvement. NICE has set examples and made advancements in involving consumers in guideline development. Feedback from consumers involved in the NICE process indicates that they value their involvement highly [28].

Oliver and colleagues identified a number of studies that evaluated different ways of involving consumers in research priority setting [31]. While the results of their systematic review are informative, there are two important limitations: 1) they focused on research priority setting and 2) direct comparison of different methods of consumer involvement were not found. A systematic review

of comparative studies of methods for involving consumers in developing health care policy, research, clinical practice guidelines and patient information found five randomised trials, but none of these were relevant to informing decisions about how best to involve consumers in developing health care recommendations [40].

Further work

We have identified a number of unresolved questions that require systematic reviews and additional research. A systematic review to evaluate the differences in values between consumers/patients and clinicians or experts is needed. Experimental work is needed that compares different strategies of consumer involvement in guideline development to evaluate whether more resource intensive approaches that include detailed methods to elicit and include values lead to different recommendations or other important differences.

A database of values assigned to specific health states may also facilitate the development of guidelines. Such a database should include information on the methods used and these methods should be explicitly stated when values are included in recommendations. While additional research on acceptable methods for eliciting values for inclusion in guidelines is required, one barrier is the philosophical and personal investment of researchers in particular methods. None of these available methods has demonstrated its superiority to others. Thus, achieving consensus on current best practice, that could be modified when new evidence become available, might be helpful.

Development of an appropriate checklist of questions that address key ethical considerations would help to ensure that these were addressed more systematically and facilitate reporting of important considerations, so that these were made more transparent. Adaptable guideline templates that allow for integration of different values should be developed and used when differences in values are likely to be critical to a decision [49]. Finally, better ways of communicating value-sensitive information need to be investigated.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

HJS prepared the first draft of this review. AF and ADO contributed to drafting and revising it. All authors read and approved the final manuscript.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper. The authors thank Mr. John Fox for administrative assistance.

References

1. Torrance GW Feeny, D: **Utilities and quality-adjusted life years.** *Int J Tech Ass Health Care* 1989, **5(4)**:559-575.
2. Cohen J: **Are clinical practice guidelines impartial.** *Int J Technol Health Care* 2004, **20**:415-420.
3. Giacomini MK Cook DJ, Streiner DL, Anand SS: **Using practice guidelines to allocate medical technologies: an ethics framework.** *Int J Technol Health Care* 2000, **16**:987-1002.
4. Hofmann B: **Toward a procedure for integrating moral issues in health technology assessment.** *Int J Technol Health Care* 2005, **21**:312-318.
5. Bennett KJ Torrance, GW: **Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques.** In *Quality of life and pharmacoeconomics in clinical trials* second edition. Edited by: Spilker B. Philadelphia, Lippincott-Raven; 1996:259.
6. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM: **Eliciting public preferences for health-care: a systematic review of techniques.** *Health Technol Assess* 2001, **5(5)**:1-186.
7. Froberg DG, Kane RL: **Methodology for measuring health-state preferences--I: Measurement strategies.** *Journal of Clinical Epidemiology* 1989, **42(4)**:345-354.
8. Brazier J Usherwood T, Harper R, Thomas K: **Deriving a preference-based single index from the UK SF-36 Health Survey.** *J Clin Epidemiol* 1998, **51**:1115-1128.
9. Torrance GW Furlong W, Feeny DH, Boyle M: **Multiattribute preferences functions: Health Utilities Index.** *Pharmacoeconomics* 1995, **6**:503-520.
10. Kaplan RM Bush JW, Berry CC: **Health Status: Types of Validity and the Index of Well-being.** *Health Services Res* 1976, **11**:478-507.
11. EuroQol-Group: **EuroQol - a new facility for the measurement of health-related quality of life.** *Health Pol* 1990, **16**:199-208.
12. Guyatt G Gutterman D, Baumann M, Addrizzo-Harris D, Hylek E, Phillips B, Raskob G, Lewis S, Schünemann H.: **Grading strength of recommendations and quality of evidence in clinical guidelines.** *Chest* 2006, **129**:174-181.
13. Buller HRA G. Hull, RD. Hyers, TM. Prins, MH. Raskob GE.: **Antithrombotic Therapy for Venous Thromboembolic Disease.** *CHEST* 2004, **126**:401S-428S.
14. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, Nagpal S, Cox JL: **Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study.** *Bmj* 2001, **323(7323)**:1218-1222.
15. Pritchard RS, Anthony SP: **Chemotherapy plus Radiotherapy Compared with Radiotherapy Alone in the Treatment of Locally Advanced, Unresectable, Non-Small-Cell Lung Cancer: A Meta-Analysis.** *Ann Intern Med* 1996, **125(9)**:723-729.
16. Robinson LA, Wagner H Jr., Ruckdeschel JC: **Treatment of Stage IIIA Non-small Cell Lung Cancer.** *Chest* 2003, **123(90010)**:202S-220.
17. **Methodology of guideline development.** *Neurosurgery* 2002, **50(3 Suppl)**:S2-6.
18. 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(1)**:18-23.

19. Hayward RS WMC Tunis SR, Bass EB, Rubin HR, Haynes RB: **More informative abstracts of articles describing clinical practice guidelines.** *Ann Intern Med* 1993, **118**:731-737.
20. Rawlins MD, Culyer AJ: **National Institute for Clinical Excellence and its value judgments.** *Bmj* 2004, **329(7459)**:224-227.
21. Fretheim A Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 3. Group composition.** *Health Res Policy Syst* 2006.
22. Oxman AD Schünemann HJ, Fretheim A.: **Improving the Use of Research Evidence in Guideline Development: 12. Incorporating considerations of equity.** *Health Res Policy Syst* 2006.
23. Schünemann HJ Fretheim A, Oxman, AD: **Improving the Use of Research Evidence in Guideline Development: 6. Determining which outcomes are important.** *Health Res Policy Syst* 2006.
24. **Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines. EIP/GPE/EQC/2003.1.** Geneva , World Health Organization; 2003.
25. **The 2000 consensus conference on clinical practice guidelines. Toronto. Health Policy Forum; 2000 Nov 2.**
26. Moynihan R Paulsen, E, Lavis, J, Oxman, AD: **A survey of producers of clinical practice guidelines and health technology assessments.** 2005.
27. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15(1)**:31-45.
28. Kelson MC: **The NICE Patient Involvement Unit.** *Evidence-Based Healthcare & Public Health* 2005, **9**:304-307.
29. Royle J, Oliver S: **Consumer involvement in the health technology assessment program.** *Int J Technol Assess Health Care* 2004, **20(4)**:493-497.
30. Oliver S: **The progress of lay involvement in the NHS Research and Development Programme.** *J Eval Clin Pract* 1996, **2(4)**:273-280.
31. Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, Gyte G, Oakley A, Stein K: **Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.** *Health Technol Assess* 2004, **8(15)**:1-148, III-IV.
32. Oliver S, Milne R, Bradburn J, Buchanan P, Kerridge L, Walley T, Gabbay J: **Involving consumers in a needs-led research programme: a pilot project.** *Health Expect* 2001, **4(1)**:18-28.
33. Brainin M, Barnes M, Baron JC, Gilhus NE, Hughes R, Selmaj K, Waldemar G: **Guidance for the preparation of neurological management guidelines by EFNS scientific task forces - revised recommendations 2004.** *European Journal of Neurology* 2004, **11(9)**:577-581.
34. Kelson MC: **Consumer collaboration, patient-defined outcomes and the preparation of Cochrane Reviews.** *Health Expectations* 1999, **2(2)**:129-135.
35. CCNET: **Consumers in Cochrane.** [<http://www.cochrane.org/consumers/homepage.htm>].
36. Oxman AD Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006.
37. **Agency for Healthcare Research and Quality** [<http://www.guidelines.gov>]
38. **Guidelines International Network** [<http://www.g-i-n.net>]
39. Dolders MGT, Zeegers MPA, Groot W, Ament A: **A Meta-analysis demonstrates no significant differences between Patient and Population Preferences.** *Journal of Clinical Epidemiology* in press.
40. Nilsen ES Myrhaug, HT, Johansen, M, Oliver, S, Oxman, AD: **Interventions for promoting consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material.** *Cochrane Library* . submitted
41. Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, Tyrer P: **Systematic review of involving patients in the planning and development of health care.** *Bmj* 2002, **325(7375)**:1263.
42. Morimoto T, Fukui T: **Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals.** *J Epidemiol* 2002, **12(2)**:160-178.
43. Owens DK: **Spine update. Patient preferences and the development of practice guidelines.** *Spine* 1998, **23(9)**:1073-1079.
44. Slomka J, Hoffman-Hogg L, Mion LC, Bair N, Bobek MB, Arroliga AC: **Influence of clinicians' values and perceptions on use of clinical practice guidelines for sedation and neuromuscular blockade in patients receiving mechanical ventilation.** *Am J Crit Care* 2000, **9(6)**:412-418.
45. Guyatt G Strauss S, McAlister F, Haynes B, Sinclair J, Devereaux PJ, Lachetti C: **Moving from Evidence to Action. Incorporating patient values.** In *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice* Edited by: Guyatt G RD. Chicago , AMA Press; 2002:567-582.
46. van Wersch A, Eccles M: **Involvement of consumers in the development of evidence based clinical guidelines: practical experiences from the North of England evidence based guideline development programme.** *Qual Saf Health Care* 2001, **10(1)**:10-16.
47. Telford R, Boote JD, Cooper CL: **What does it mean to involve consumers successfully in NHS research? A consensus study.** *Health Expect* 2004, **7(3)**:209-220.
48. Schünemann HJ, Fretheim A, Oxman AD: **Templates, adaptation and the role of international guidelines.** . (submitted as part of this series)
49. Schünemann HJ Oxman AD, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 13. Adaptation, applicability and transferability.** *Health Res Policy Syst* 2006.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications

Tessa Tan-Torres Edejer*

Address: Health Systems Financing, World Health Organization, CH-1211 Geneva 27, Switzerland

Email: Tessa Tan-Torres Edejer* - tantorrest@who.int

* Corresponding author

Published: 05 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:23 doi:10.1186/1478-4505-4-23

Accepted: 05 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/23>

© 2006 Edejer; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 11th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on incorporating considerations of cost-effectiveness, affordability and resource implications in guidelines and recommendations.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: When is it important to incorporate cost-effectiveness, resource implications and affordability considerations in WHO guidelines (which topics)?

- For cost-effectiveness:

The need for cost/effectiveness information should be dictated by the specific question, of which several may be addressed in a single guideline. It is proposed that the indications for undertaking a cost-effectiveness analysis (CEA) could be a starting point for determining which recommendation(s) in the guideline would benefit from such analysis.

- For resource implications/affordability:

The resource implications of *each* individual recommendation need to be considered when implementation issues are being discussed.

How can cost-effectiveness, resource implications and affordability be explicitly taken into account in WHO guidelines?

- For cost-effectiveness:

- If data are available, the ideal time to consider cost-effectiveness is during the evidence gathering and synthesizing stage. However, because of the inconsistent availability of CEAs and the procedural difficulty associated with adjusting results from different CEAs to make them comparable, it is also possible for cost-effectiveness to be considered during the stage of developing recommendations.

- Depending on the quantity and quality and relevance of the data available, such data can be considered in a qualitative way or in a quantitative way, ranging from a listing of the costs to a modelling exercise. At the very least, a qualitative approach like a commentary outlining the economic issues that need to be considered is necessary. If a quantitative approach is to be used, the full model should be transparent and comprehensive.

- For resource implications/affordability:
 - Resource implications, including health system changes, for each recommendation in a WHO guideline should be explored. At the minimum, a qualitative description that can serve as a gross indicator of the amount of resources needed, relative to current practice, should be provided.

How does one provide guidance in contextualizing guideline recommendations at the country level based on considerations of cost-effectiveness, resource implications and affordability?

- All models should be made available and ideally are designed to allow for analysts to make changes in key parameters and reapply results in their own country.
- In the global guidelines, scenarios and extensive sensitivity/uncertainty analysis can be applied.

Resource implications for WHO

- From the above, it is clear that guidelines development groups will need a health economist. There is need to ensure that this is included in the budget for guidelines and that there is in-house support for this as well.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 11th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this. In this paper we address the following questions:

- When is it important to incorporate cost-effectiveness, resource implications and affordability considerations in WHO guidelines (which topics)?
- How can cost-effectiveness, resource implications and affordability be explicitly taken into account in WHO guidelines?
- How does one provide guidance in contextualizing guideline recommendations at the country level based on considerations of cost-effectiveness, resource implications and affordability?
- What are the resource implications of the answers to these questions for WHO?

What is WHO doing now?

In 2003, the Guidelines for WHO Guidelines [1] recommended that both cost-effectiveness and resource implications of guideline recommendations be considered when developing WHO guidelines. With the primary audience being Ministry of Health officials with a mandate to improve population health rather than the health of individuals, such concerns were considered to be appropriate. The guidelines (see Additional file 1 for relevant portions) states very briefly that cost-effectiveness is to be considered during the stage of formulation of the recommenda-

tions and that the recommendations could be expressed in terms of scenarios (for countries with limited versus some resources). It also recommends that 1–2 experts in cost-effectiveness analysis (CEA) be involved in the technical guidelines development group. Finally, in the third stage or the localization phase in-country, it suggests that WHO provide technical assistance for countries to use their own or regional cost-effectiveness information and that WHO also provide information on financial costs of implementing the recommendations.

There has been some mention of cost-effectiveness in a few guidelines that were issued since 2003 and that claim to have used the Guidelines for WHO Guidelines. The 2003 WHO/ISH guidelines [2] had general sections on cost-effectiveness and resource implications of the recommendations. More recently, the WHO "guidelines" on hand hygiene in health care [3] also included information on economic burden of poor hygiene, cost-effectiveness and resource implications of recommendations. Another example, the malaria treatment guidelines [4] specifically state that cost-effectiveness studies were not included for consideration because there were very few completed studies relating to the interventions being considered, and that the prices of the anti-malarial drugs were fluid, "rendering such studies unreliable."

What other organizations are doing?

It is clear that cost-effectiveness and/or cost implications of recommendations are recognized in guideline development. The US Preventive Services Task Force [5] lists several reasons why CEA is useful in guideline development:

1. Quantifying the differences between two or more effective services for the same condition

2. Illustrating the impact of delivering a given intervention at different intervals, different ages, or to different risk groups
3. Evaluating the potential role of new technologies
4. Identifying key conditions that must be met to achieve the intended benefit of an intervention
5. Incorporating preferences for intervention outcomes
6. Developing a ranking of services in order of their costs and expected benefits

The AGREE guidelines appraisal instrument [6], on which the WHO checklist was loosely based, includes the cost impact of guideline recommendations under their applicability criteria.

To actually gauge what other organizations are doing with respect to cost-effectiveness/costs in guideline development, one can review the guidelines that have been issued by these organizations and/or review their documented methods for guideline development. It is hoped that there is consistency between the two. In 1999, 279 guidelines that were published in peer-reviewed literature were reviewed [7]. Only 41.6% made any mention of projected effects on health care costs, and only 14.3% quantified these estimates in any way. A 2002 study on the incorporation of published cost-effectiveness analysis in published clinical guidelines showed that, using guidelines as the unit of analysis, 9 of 35 (26%) incorporated at least 1 economic analysis of above-average quality in the text and 11 of 35 (31%) incorporated at least 1 in the references [8]. Finally, a search of the database of the National Guideline Clearinghouse [9] showed that of 1616 guidelines published between 2000–2005, only 369 or 23% had a formal cost analysis.

For a review of methods, a 2003 survey [10] of 18 clinical practice guideline development agencies showed that six included costs/cost containment/cost-effectiveness in their objectives but only three routinely included health economists in their guideline development groups (NHRMC in Australia, SBU in Sweden, and North of England). The AGREE prototype electronic library shows some illustrative excerpts specifically mentioning costs from the methodologies of five national agencies. More detailed information on the use of costs/cost-effectiveness information by guidelines agencies is available (see Additional file 2).

To what extent these organizations routinely and explicitly use costs/CEA in their guidelines is not clear. More information is needed on the actual experience of these

organizations in incorporating costs/cost-effectiveness information during their guideline development process (e.g. lessons learned) [11] and an assessment of whether economic evaluations have provided added value to their guidelines [12].

Methods

The methods used to prepare this review are described in the introduction to this series [13]. Briefly, the key questions addressed in this paper were vetted amongst the authors of the series of articles and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct full systematic reviews. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register, the US National Guideline Clearinghouse, and the Guidelines International Network for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review PubMed was searched using the following text word searches: costs and generalizability, practice guidelines and cost-effectiveness analysis and combinations thereof. Using the same search words, the Internet search machine, Google, was also used to search for unpublished documents. Websites of known repositories of guidelines, of organizations of guideline agencies, and of pioneer/well-known guideline development agencies were visited. References in key documents that had titles which could be probably relevant were also pursued. A few times, when it was evident that an author had this as his/her special area of interest, a search using his/her name in connection with guidelines was also done. Papers were included if they described guidelines that included cost-effectiveness information, or described methods of guideline agencies to incorporate costs/resource implications/CEA information.

The information collected during the review was then synthesized where relevant for each question and was used as the basis to draw the implications for WHO guidelines.

Findings

When is it important to incorporate considerations of cost-effectiveness and resource implications of recommendations in WHO guidelines (which topics)?

The need for cost-effectiveness information should be dictated by the specific question of which several may be addressed in one guideline alone. It is proposed that the indications for undertaking a cost-effectiveness analysis could be a starting point for determining which recommendation(s) in the guideline would benefit from such analysis [14]. From the review, NHRMC states explicitly

that "the challenge is to focus on the decision points that are of key importance in an economic sense and pinpoint the nature of economic information needed to address these questions. The key decisions concern health care that contributes significantly to the total cost of an option, options with very different costs or care that contributes significantly to health outcomes. On the other hand, decisions are unimportant if they concern health care that is uncontroversial, options that are not economically viable or options for which there are no large resource implications." [15]

The UK National Institute for Health and Clinical Excellence (NICE) extends this concept by adopting a value of analysis approach where aside from "the overall 'importance' of the recommendation (which is a function of the number of patients affected and the potential impact on costs and health outcomes per patient)", they also suggest evaluating "the current extent of uncertainty over cost-effectiveness and the likelihood that analysis will reduce this uncertainty" [16].

Implications for WHO guidelines

At the scoping stage, the assistance of an experienced health economist who is familiar with the area of interest would be needed and a selective identification of the issues needing CEA could be done. If CEA is not initially identified during the scoping stage as a clear need, this issue needs to be revisited again at the evidence, recommendations and peer review stage.

On the other hand, the resource implications of *each* individual recommendation needs to be considered when implementation issues are being discussed. This may be done at the global guideline level, through the use of scenarios, and at the local adaptation or country level.

How can cost-effectiveness and affordability be explicitly taken into account in WHO guidelines?

There are different points in the guideline development process that cost-effectiveness and resource implications of guidelines recommendations can be considered. The first phase is the evidence phase and to the extent possible, this is where cost-effectiveness information should be considered.

As in questions on effectiveness, the question can be raised as to why conduct a review of CEAs rather than identify a single study that addresses the question? "CEAs vary widely in their methods and assumptions. Because of this variation, systematically reviewing CEAs provides several benefits. First, because CEAs draw on a variety of cost and effectiveness data sources to develop input parameters, a systematic review can identify which analyses use the best available evidence for key inputs and are therefore

the most evidence based. Second, because the credibility of CEAs rests on their quality, a critical review of CEAs and a rating of the quality of each allow for identifying the most methodologically rigorous studies. Third, a comprehensive review can identify the studies that best address the question being asked. Fourth, comparatively assessing CEAs can help to identify variables and methods that significantly influence the estimated benefits and cost effectiveness of an intervention. For instance, some CEAs might assume no harms from a given intervention, while others might assume that the intervention has significant harms. Comparing these studies side by side may provide insight into how the assumption or lack of assumption of harm affects the estimated benefit of the intervention. While some assumptions are varied within a single study using sensitivity analysis, most CEAs provide a limited number of sensitivity analyses. Thus, systematically reviewing CEAs may help identify, through a side-by-side comparison that amounts to a "virtual sensitivity analysis," the impact of different assumptions on the benefits of a given intervention. Finally, the more high-quality, independently conducted CEAs there are for a given intervention, the more convincing the evidence." [17]

In doing a systematic review of economic evaluations, the first step is to search for the literature. There are available resources on the internet which list databases and compilations of economic evaluations [18]. In addition, there has been a systematic evaluation in terms of sensitivity and specificity of different search strategies for economic evaluations [19].

In extracting data, considerable progress has been reported with development of a systematic process of adjustments of results from different studies to make them comparable. The Task Force on Community Preventive Services admits that "no process of adjustment or other means of reviewing existing economic evaluations is flawless." But it makes the point that to adjust data to make it comparable is better than to "(1) ignore economic information entirely; (2) attempt to use non-comparable data; or (3) adjust in ways that are not systematic or explicit" [20].

Some work has also been done in terms of very simple visual methods to present summaries of cost-effectiveness analysis [21]. Despite all of these advances methodologically, the current situation shows however, that there is limited availability and variable quality of relevant CEAs [22].

Implications for WHO guidelines

If data is available, the ideal time to consider CEA is during the evidence gathering and synthesizing stage. However, because of the inconsistent availability of CEAs and

the procedural difficulty associated with adjusting results from different CEAs to make them comparable [20] it is also possible for CEA to be considered during the stage of developing recommendations. This is also consistent with the GRADE approach [23]. At this stage, the information of the resource implications and outcomes of the recommended interventions can be considered simultaneously. Depending on the quantity and quality and relevance of the data available, such data can be considered in a qualitative way or in a quantitative way, ranging from a listing of the costs to a modelling exercise [24,25]. At the very least, a qualitative approach like a commentary outlining the economic issues that need to be considered is necessary [26]. If a quantitative approach is to be used, the full model should be transparent, be made available and extensive uncertainty/sensitivity analysis built-in so as to allow analysts to selectively reapply results in their own country, as in the WHO-CHOICE contextualization tool [27].

How does one provide guidance in contextualizing guideline recommendations at the country level based on considerations of cost-effectiveness and affordability?

For cost-effectiveness, there are concerns about the generalizability of results from a single CEA or even a systematic review of a CEA. A review of sources of variability frequently mentions volume and costs of resources consumed as a source of variability [28]. Not as much work has been done on variability of outcomes. Very recently, a checklist was developed for assessing variability or generalizability to be able to translate information from one developed country to another [29]. There is a need to pilot test this, a revision or another instrument in developing countries. For costs, more specifically prices, general principles for adaptation are available [30].

Affordability or resource implications can be considered in the global guidelines if it gives guidance by provision of basic information that will allow guideline users to work out the cost implications for their own service [25,15]. A scenario approach can be used. Also, in this context, WHO-CHOICE data and methods are useful for contextualization [31]. Note that this exercise will also need to include the health system implications of the recommendations, from training, changes in supervision, monitoring and evaluation, advocacy, etc. as seen in some recent examples [32,33].

Implications for WHO

Resource implications, including health system changes, for each recommendation in WHO guidelines should be explored. At the minimum, a qualitative description that can serve as a gross indicator of the amount of resources needed, relative to current practice, should be provided.

Overall assessment of need for health economics expertise

In summary, the role of the health economist in a guideline development group is to:

- help to identify the clinical issues or questions for economic analysis
- review economic literature
- carry out or commission cost-effectiveness analyses
- estimate the cost and resource implications of the recommendations.

"The relative weight given to each role will vary from guideline to guideline. There may be large differences between guidelines in respect of the literature available to review: the size of the relevant economics literature, its relevance, its quality, its timeliness, its generalisability. In some areas there may be good-quality data that can be used in economic models, whereas other areas may have a dearth of such data." [16] Additionally, all throughout the process, the health economist can educate the other guideline development group about CEA and vice-versa, through interaction with the guideline development group members, s/he will be better able to work with an improved understanding of the health issues being considered.

Implications for WHO

From the above, it is clear that guidelines development groups will need access to a health economist. There is need to ensure that this is budgeted for in the resources and that there is in-house support for this as well.

Further work

There is a need to: 1) get more information on the actual experience of guideline agencies in incorporating CEA in guidelines; 2) assess the added value of economic evaluations in guidelines by comparing recommendations with and without CEA; 3) further expand the section on CEA and resource implications in guidelines, including specification of the minimum information that should be provided, in collaboration with health economists and experienced guideline developers.

Competing interests

The author helped to develop the current Guidelines for WHO Guidelines.

Authors' contributions

TTTE prepared the first draft of the review and was responsible for the subsequent revisions.

Additional material

Additional file 1

Cost-effectiveness and affordability in the Guidelines for WHO Guidelines. Excerpts from the document "Guidelines for WHO Guidelines" which refer to the concepts of cost-effectiveness and affordability.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1478-4505-4-23-S1.doc>]

Additional file 2

Considerations of costs/cost-effectiveness information by other agencies. Excerpts from documents from other agencies that issue guidelines which refer to the concepts of costs/cost-effectiveness

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1478-4505-4-23-S2.doc>]

Acknowledgements

I gratefully acknowledge the members of the ACHR SURE, and all of those who have provided feedback on earlier versions of this paper.

References

- Guidelines for WHO guidelines: World Health Organization. *Global Programme on Evidence for Health Policy* [http://whqlibdoc.who.int/hq/2003/EIP_GPE_EOC_2003_1.pdf].
- 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) guidelines statement on management of hypertension 2003 [http://www.who.int/cardiovascular_diseases/guidelines/hypertension_guidelines.pdf]. Lip-pincott, Williams and Wilkins
- WHO guidelines on hand hygiene in health care (advanced draft) : a summary : clean hands are safer hands 2005 [http://whqlibdoc.who.int/hq/2005/WHO_EIP_SPO_OPS_05.2.pdf]. World Health Organization. Dept. of Health System Policies and Operations
- World Health Organization: *Guidelines for the Treatment of Malaria* Geneva, World Health Organization; 2006.
- Saha S, Hoerger TJ, Pignone MP, Teutsch SM, Helfand M, Mandelblatt JS, Cost Work Group, Third US Preventive Services Task Force: **The art and science of incorporating cost effectiveness into evidence-based recommendations for clinical preventive services.** *Am J Prev Med* 2001, **20(3 Suppl)**:36-43.
- The AGREE Collaboration: **Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument.** 2001 [<http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf>].
- 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* **281(20)**:1900-5. 1999 May 26
- Wallace JF, Weingarten SR, Chiou CF, Henning JM, Hohlbauch AA, Richards MS, Herzog NS, Lewensztain LS, Ofman JJ: **The limited incorporation of economic analyses in clinical practice guidelines.** *J Gen Intern Med* 2002, **17(3)**:210-20.
- National Guidelines Clearinghouse [<http://www.guideline.gov>]
- 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(1)**:31-45.
- Louis W: **Niessen, Els Grijseels, Marc Koopmanschap and Frans Rutten. Economic analysis for clinical practice – the case of 31 national consensus guidelines in the Netherlands.** *Journal of Evaluation in Clinical Practice* (forthcoming) .
- Granata AV, Hillman AL: **Competing practice guidelines: using cost-effectiveness analysis to make optimal decisions.** *Ann Intern Med* **128(1)**:56-63. 1998 Jan 1
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Drummond Michael F, O'Brien Bernie, Stoddart Greg L, Torrance George W: *Methods for the Economic Evaluation of Health Care Programmes* Second edition. Oxford University Press; 1997.
- NHRMC: **How to compare the costs and benefits: evaluation of the economic evidence Handbook series on preparing clinical practice guidelines; Endorsed July 2001.** [<http://www7.health.gov.au/nhmrc/publications/synopses/cp73syn.htm>].
- National Institute of Clinical Excellence: **The Guidelines Manual. Incorporating health economics in guidelines and assessing resource impact Chapter 8:** [<http://www.nice.org.uk/page.aspx?o=308777>]. 25 April 2006.
- Truman BI, Smith-Akin CK, Hinman AR, Gebbie KM, Brownson R, Novick LF, Lawrence RS, Pappaioanou M, Fielding J, Evans CA Jr, Guerra FA, Vogel-Taylor M, Mahan CS, Fullilove M, Zaza S: **Developing the Guide to Community Preventive Services – overview and rationale. The Task Force on Community Preventive Services.** *Am J Prev Med* 2000, **18(1 Suppl)**:18-26.
- NHS Economic Evaluation Database** [<http://www.york.ac.uk/inst/crd/econ4.htm>]
- Sassi F, Archard L, McDaid D: **Searching literature databases for health care economic evaluations: how systematic can we afford to be?** *Med Care* 2002, **40(5)**:387-94.
- Carande-Kulis VG, Maciosek MV, Briss PA, Teutsch SM, Zaza S, Truman BI, Messonnier ML, Pappaioanou M, Harris JR, Fielding J: **Methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services. Task Force on Community Preventive Services.** *Am J Prev Med* 2000, **18(1 Suppl)**:75-91.
- Nixon J, Khan KS, Kleijnen J: **Summarising economic evaluations in systematic reviews: a new approach.** *BMJ* **322(7302)**:1596-8. 2001 Jun 30
- Jefferson T, Demicheli V, Vale L: **Quality of systematic reviews of economic evaluations in health care.** *JAMA* **287(21)**:2809-12. 2002 Jun 5
- 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, Schunemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S, GRADE Working Group: **Grading quality of evidence and strength of recommendations.** *BMJ* **328(7454)**:1490. 2004 Jun 19
- Eccles M, Mason J: **How to develop cost-conscious guidelines.** *Health Technol Assess* 2001, **5(16)**:1-69.
- Phillips Z, Ginelly L, Sculpher M, et al.: **Review of good practice in decision-analytic modelling in health technology assessment.** *Health Technology Assessment* 2004, **8(36)**:.
SIGN 50: A guideline developers' handbook [<http://www.sign.ac.uk/guidelines/fulltext/50/section6.html>]
- CHOosing Interventions that are Cost Effective (WHO-CHOICE)** [<http://www.who.int/choice>]
- Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A: **Generalisability in economic evaluation studies in healthcare: a review and case studies.** *Health Technol Assess* 2004, **8(49)**:iii-iv. 1-192.
- Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, de Pouvourville G: **Can economic evaluations be made more transferable?** *Eur J Health Econ* in press. 2005 Oct 26
- Hutton G, Baltussen R: **Cost valuation in resource-poor settings.** *Health Policy Plan* 2005, **20(4)**:252-9.
- Hutubessy R, Chisholm D, Edejer TT: **Generalized cost-effectiveness analysis for national-level priority-setting in the health sector.** *Cost Eff Resour Alloc* **1(1)**:8. 2003 Dec 19
- Deen JL, Weber M, Qazi S, Fontaine O: **An extended field test of the WHO severe malnutrition guidelines in hospitals in developing countries: preliminary observations.** *Forum Nutr* 2003, **56**:183-4.
- Parati G, Mendis S, Abegunde D, Asmar R, Mieke S, Murray A, Shengelia B, Steenvoorden G, Van Montfrans G, O'Brien E: **Recommendations for blood pressure measuring devices for office/clinic use in low resource settings.** *Blood Press Monit* 2005, **10(1)**:3-10.

Review

Improving the use of research evidence in guideline development: 12. Incorporating considerations of equity

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim¹

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 05 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:24 doi:10.1186/1478-4505-4-24

Accepted: 05 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/24>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 12th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on incorporating considerations of equity in guidelines and recommendations.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We found few directly relevant empirical methodological studies. These answers are based largely on logical arguments.

When and how should inequities be addressed in systematic reviews that are used as background documents for recommendations?

- The following question should routinely be considered: Are there plausible reasons for anticipating differential relative effects across disadvantaged and advantaged populations?
- If there are plausible reasons for anticipating differential effects, additional evidence should be included in a review to inform judgments about the likelihood of differential effects.

What questions about equity should routinely be addressed by those making recommendations on behalf of WHO?

- The following additional questions should routinely be considered:
- How likely is it that the results of available research are applicable to disadvantaged populations and settings?
- How likely are differences in baseline risk that would result in differential absolute effects across disadvantaged and advantaged populations?
- How likely is it that there are important differences in trade-offs between the expected benefits and harms across disadvantaged and advantaged populations?
- Are there different implications for disadvantaged and advantaged populations, or implications for addressing inequities?

What context specific information is needed to inform adaptation and decision making in a specific setting with regard to impacts on equity?

- Those making recommendations on behalf of WHO should routinely consider and offer advice about the importance of the following types of context specific data that might be needed to inform adaptation and decision making in a specific setting:
- Effect modifiers for disadvantaged populations and for the likelihood of differential effects
- Baseline risk in relationship to social and economic status
- Utilization and access to care in relationship to social and economic status
- Costs in relationship to social and economic status

- Ethics and laws that may impact on strategies for addressing inequities
- Availability of resources to address inequities

What implementation strategies are likely be needed to ensure that recommendations are implemented equitably?

- Organisational changes are likely to be important to address inequities. While it may only be possible to consider these in relationship to specific settings, consideration should be given to how best to provide support for identifying and addressing needs for organisational changes. In countries with pervasive inequities institutional, cultural and political changes may first be needed.
 - Appropriate indicators of social and economic status should be used to monitor the effects of implementing recommendations on disadvantaged populations and on changes in social and economic status.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 12th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Braveman and Gruskin define equity as "the absence of disparities in health that are systematically associated with social advantage or disadvantage" [1]. The message is made clearer by Margaret Whitehead's definition of inequity: "differences in health which are not only unnecessary and avoidable but, in addition, are considered unfair and unjust" [2]. Inequities in health and health care are well documented in relationship to social and economic factors, including Place of residence (e.g. rural, urban, inner city, Race/ethnicity/culture, Occupation, Gender, Religion, Educational level, Socioeconomic status and Social capital (availability of neighbourhood support, social stigma, civic society) (PROGRESS) [3].

Disadvantaged populations almost always have poorer health [4], poorer access to health care [5], and receive poorer quality health care [6]. To the extent that recommendations influence what is done, they can improve the overall health of the population but have no impact on inequities, reduce inequities or exacerbate them regardless of the overall effects on population health. There has been a growing interest in taking equity into consideration in clinical practice guidelines [7,8]. However, consideration of inequities has generally been lacking [7]. For example, AGREE and other instruments for assessing the quality of guidelines do not include items on equity or the fairness of the recommendations [9].

In this paper we address the following questions:

- When and how should inequities be addressed in systematic reviews that are used as background documents for recommendations?

- What questions about equity should routinely be addressed by those making recommendations on behalf of WHO?

- What context specific information is needed to inform adaptation and decision making in a specific setting with regard to impacts on equity?

Related questions about adaptation, applicability and transferability are addressed in another paper in this series [10].

What WHO is doing now

"WHO has embraced the elimination of health inequities as an important target and supports the dual goals of equity and efficiency for health services. WHO's data gathering on inequalities in health status and access to services is shaped by and in turn informs its advocacy and normative activities that aim to reduce health inequities. Besides collecting relevant data broken down by group, WHO attempts both to relate these data to health determinants (e.g., membership in less privileged social groups and exposure to various hazards) and to develop and disseminate interventions to improve conditions for members of such groups" [11].

Nonetheless, we are not aware of any specific documents that provide guidance as to how equity should be taken into account in WHO guidelines or recommendations or of any studies or descriptions of current practice. The WHO guidelines for guidelines do not currently provide any explicit advice regarding how to take account of equity.

What other organisations are doing

Clinical practice guidelines typically focus on the effectiveness of interventions (Will adherence to a recommendation do more good than harm?), occasionally on cost-effectiveness (Are the net benefits worth the costs?), and rarely on equity (Are the recommendations fair?) [7]. More recently, several guideline developers have begun to consider equity explicitly and systematically, including, for example, the Australian NHMRC [7], INCLIN [8], the GRADE Working Group, and the National Institute for

Health and Clinical Excellence (NICE) in the UK, which now has an extended mandate including public health guidance and reducing health inequalities, after the Health Development Agency (HDA) became part of NICE in 2005 [12]. The HDA was established in 2000 to develop the evidence base to improve health and reduce health inequalities. It worked in partnership with professionals and practitioners across a range of sectors to translate that evidence into practice. Other countries that have had a major political commitment to reducing inequities in health include the Netherlands [13], Thailand, and Chile [14].

Methods

The methods used to prepare this review are described in the introduction to this series [15]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register [16], the US National Guideline Clearinghouse [17], and the Guidelines International Network [18]) for existing systematic reviews and relevant methodological research that address these questions. We did not conduct systematic reviews ourselves. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

This paper is based in large part on a workshop on addressing inequities held in Oslo August 31 to September 1, 2005 [19], background documentation for that workshop [20-23], and a reference list generated during and subsequent to the workshop. We searched PubMed using (clinical practice guidelines or public health guidelines) and (equity or equality) and related articles for references [7] and [23]. We searched the Cochrane Methodology Register using equity or equality.

Findings

Our database searches yielded few references and we found few directly relevant empirical methodological studies, consistent with the findings of other reviews [22,23]. For example, the literature search and correspondence with guideline developers worldwide by the NHMRC located no examples of where clinical practice guideline developers explicitly incorporated evidence on socioeconomic position and health into generic guidelines, except for when guidelines were developed for specific disadvantaged sub-populations [22]. This is consistent with the findings of the Health Development Agency in England. They observed that there is a very large literature that describes the problem of inequalities and a

very much smaller one describing interventions that could reduce inequalities [24].

When and how should inequities be addressed in systematic reviews that are used as background documents for recommendations?

Evidence of the effects of interventions on inequities is sparse and difficult to search for [25]. For example, Tsikata and colleagues found that only 10% of controlled trials assessed the efficacy of the intervention across socioeconomic subgroups [26]. Similarly, Ogilvie and colleagues found that in Cochrane reviews of controlled studies of tobacco control both the reviews and the primary studies in those reviews rarely assessed the impact of the intervention across socioeconomic factors [27]. Systematic reviews tend not to provide evidence on differential effectiveness [27-33]. Searches of electronic databases in many fields, particularly for social interventions and more upstream interventions, may miss much relevant evidence [31-33]. Publication bias may be a problem [25]. Because there is limited direct evidence of differential effects of interventions across socioeconomic groups, it will generally be necessary to search for and include a wider scope of evidence to support or refute plausible hypotheses of differential effects, or the effects of interventions on reducing inequities.

Although there are clear arguments for exploring moderator effects in systematic reviews, subgroup analyses can be misleading both because of inadequate power (resulting in false negative conclusions) and multiple testing (resulting in false positive conclusions) [34-38]. The results observed in subgroups may differ by chance from the overall effect identified by the meta-analysis, and the subgroup findings may not be confirmed by subsequent large trials [36,39]. Paradoxically, the best estimate of the outcome of the intervention in a sub-group may come from discounting the results of the sub-group analysis and using the overall results (Stein's paradox) [36,40]. General guidelines for interpreting subgroup analyses can be applied to subgroup analyses based on socioeconomic factors [40,41].

What questions about equity should routinely be addressed by those making recommendations on behalf of WHO?

Additional questions that should be considered in relationship to equity include questions about the applicability of the evidence to disadvantaged populations, differences in values, and the implications of these differences. General guidelines for considering the applicability of evidence can be applied to considering the applicability of evidence to disadvantaged populations [42], including differences in absolute effects due to differences in baseline risk. The trade-offs between the benefits and harms of

an intervention may be different because of differences in the relative or absolute effects of an intervention or because of differences in values [8]. For example, if an outcome, such as the ability to quickly return to or stay at work, is more important to disadvantaged populations, this might tip the balance between the benefits, harms and costs of an intervention (for example antiretrovirals for AIDS) in favour of intervening. Differences in any of these factors can result in different implications and recommendations for disadvantaged populations or specific recommendations for addressing inequities [8].

What context specific information is needed to inform adaptation and decision making in a specific setting with regard to impacts on equity?

While evidence about the effects of interventions generally comes from global research, it is necessary to take into account factors in a specific setting to inform decisions about what to do. These factors include each of the following in relationship to socioeconomic factors: the presence of effect modifiers that have been identified in the global research, baseline risk, utilization and access to care, and costs. In addition, it is necessary to take into account relevant ethical and legal standards in a specific setting, and the availability of resources to address inequities. Although this information is beyond the scope of a review or international guidelines or recommendations, international groups can systematically consider the need for these different types of information in specific settings and provide guidance regarding the importance of obtaining such information and practical strategies for doing so and integrating context specific information into decision-making processes.

What implementation strategies are likely be needed to ensure that recommendations are implemented equitably?

Because disadvantaged populations generally have poorer access to care and often receive poorer quality care, organisational changes are likely to be needed to address inequities in health care. Organisational changes are also likely to be necessary to implement interventions targeted at social determinants of health. Identifying necessary organisational changes, and barriers and facilitators of implementing change requires context specific knowledge and decisions. Nonetheless, general guidance and support for what information to consider, possible strategies to address common barriers and facilitators, and general frameworks for planning organisational changes and implementation strategies can be provided internationally. In countries with pervasive inequities institutional, cultural and political changes may first be needed.

Similarly, although local data are needed to monitor the effects of implementing recommendations, guidance can be provided regarding appropriate indicators of social gra-

dients and measures of change (e.g. in the ratio of quintile 1 to 5, or concentration indices) to use in order to monitor the effects of implementing recommendations on disadvantaged populations and on changes in social gradients. Because the evidence for interventions to reduce inequities will commonly be weak, it is generally important to ensure that monitoring and evaluations are as rigorous as possible to ensure that intended effects are achieved and unintended adverse effects are avoided.

What 'maps' are available of the different dimensions of inequity locally?

Equity and inequity are not one-dimensional phenomena. They consist of a number of dimensions that include economic status, occupation, gender, ethnicity, class, caste, religion, status grouping, age, disability, place of residence, geographical location, and manifest sexual orientation. These different dimensions are of varying salience in any given social context. For example caste and religion are more frequently significant in pre industrial systems while occupation tends to be dominant in industrial systems. It is also important to note that the importance of these various dimensions relative to each other also varies, as the dimensions overlap and overlay each other. The health effects of inequities are a product of the interplay of these different dimensions. It is therefore important to describe systematically the dimensions, and if possible their relative salience, in any given social arrangement.

Discussion

Inequities are rarely addressed in clinical practice guidelines. Evidence of the effects of public health and health policy interventions on reducing inequities is generally weak or lacking [43]. As a consequence, advice regarding how to address inequities in recommendations must to a large extent rely on the application of general methodological studies and principles, for example in relationship to subgroup analyses and applicability. While addressing inequities is a fundamental concern at the heart of WHO's mission, at present there appears to be inadequate guidance on how best to do this in developing and implementing recommendations.

Although we have not found empirical descriptions of WHO's current practices, it is reasonable to assume that inequities are not being addressed systematically and transparently. This assumption rests in part on documentation that WHO guidelines generally have not adhered to standards such as AGREE [44,45]. WHO may be more likely to address inequities than many other organisations, given its mission. However, the available evidence suggests that inequities are generally not well addressed in most systematic reviews and clinical practice guidelines. It is only recently that attention has been given to the meth-

ods used to address inequities, both for clinical and public health interventions [7,21,46].

Further work

Although we have not conducted a systematic review of the relevant literature, a more systematic review is not likely to have results or implications that are substantially different, given the sparseness of methodological research in this area. This assumption is supported by the NHMRC review [7] and a NHS HTA review of addressing equity in economic analyses [23]. However, growing attention is being paid to this area and there are areas of research that can further inform specific issues, such as the selection of indicators of socioeconomic status in relationship to specific interventions or conditions. Thus, while we do not believe that WHO should undertake further work at this time, it would be valuable for WHO or others to undertake and keep up-to-date systematic methodology reviews that address specific aspects of how to address inequities in systematic reviews, guidelines and recommendations.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We would like to acknowledge the participants in the workshop on addressing inequities held in Oslo, August 31 to September 1, 2005, who contributed to much of the thinking reflected in this review, and the Norwegian Directorate for Health and Social Welfare, which funded the workshop. Participants in the workshop included Inday Dans, Tony Dans, Mike Kelly, Sally McIntyre, Trevor Sheldon, Peter Tugwell. We would also like to thank Mark Petticrew and Elizabeth Waters for their contributions. We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Braveman PA, Gruskin S: **Defining equity in health.** *J Epidemiol Community Health* 2003, **57**:254-8.
- Whitehead M: **The concepts and principles of equity in health.** *Int J Health Serv* 1992, **22**:429-45.
- World Health Organization: **World Report on Knowledge for Better Health.** Geneva: WHO; 2004:114.
- Marmot M: **Social determinants of health inequalities.** *Lancet* 2005, **365**:1099-104.
- Wilkinson R, Marmot M, ed: **Social determinants of health.** In *The solid facts* Copenhagen:WHO; 1998:308.
- Kahn Katherine L, Marjorie Pearson, Harrison Ellen R, Rogers William H, Brook Robert H, Katherine Desmond, Keeler Emmett B: **Analysis of Quality of Care for Patients Who Are Black or Poor in Rural and Urban Settings.** *RAND Document* 1993, xviii:139.
- Aldrich R, Kemp L, Williams JS, Harris E, Simpson S, Wilson A, McGill K, Byles J, Lowe J, Jackson T: **Using socioeconomic evidence in clinical practice guidelines.** *BMJ* 2003, **327**:1283-5.
- Dans AM, Dans LF, Oxman AD, Robinson V, Acuin J, Tugwell P, Dennis R, Kang D: **Addressing inequities in clinical practice guidelines.** 2003 in press.
- The AGREE Collaboration** [<http://www.agreecollaboration.org/>]
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 13. Adaptation, applicability and transferability.** *Health Res Policy Syst* 2006.
- World Health Organisation. Equity** [<http://www.who.int/trade/glossary/story024/en/print.html>]
- Littlejohns P, Kelly M: **The changing face of NICE: the same but different.** *Lancet* 2005, **366**:791-4.
- Mackenbach JP, Stronks K: **A strategy for tackling health inequalities in the Netherlands.** *BMJ* 2002, **325**:1029-32.
- Moynihan R: **Using Health Research in Policy and Practice: Case Studies from Nine Countries.** New York: Milbank Memorial Fund; 2004.
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- National Guideline Clearinghouse** [<http://www.guidelines.gov>]
- Guidelines International Network (GIN)** [<http://www.g-i-n.net>]
- Dans AM, Dans LF, Kelly M, McIntyre S, Oxman AD, Sheldon T, Tugwell P: **Oslo workshop on addressing equity in systematic reviews, health technology Assessments, Clinical Practice Guidelines and policy briefs. Unpublished report.**
- Methodological Problems in Constructing the Evidence Base in Public Health (Mike Kelly, Catherine Swann, Amanda Killoran, Bhash Naidoo, Elaine Barnett-Paige, & Antony Morgan. Health Development Agency Public Health Evidence Steering Group Methodology Reference Group. Starter Paper. 9 August 2002.**
- Systematic Reviews of Health Promotion and Public Health Interventions (Nicki Jackson, Elizabeth Waters et al. The Cochrane Health Promotion and Public Health Field. February 2005.)**
- Using socioeconomic evidence in clinical practice guidelines (NHMRC 2002)** [<http://www.nhmrc.gov.au/publications/synopses/cp89syn.htm>]
- Sassi F, Archard L, Le Grand J: **Equity and the economic evaluation of healthcare.** *Health Technol Assess* 2001, **5(3)**:1-138.
- Millward L, Kelly MP, Nutbeam D: **Public Health Interventions Research: The Evidence.** 2003 [<http://www.publihealth.nice.org.uk/page.aspx?o=502583>]. London: Health Development Agency
- Howes F, Doyle J, Jackson N, Waters E: **Evidence-based public health: The importance of finding 'difficult to locate' public health and health promotion intervention studies for systematic reviews.** *J Epidemiol Community Health* 2004, **26**:101-4.
- Tsikata S, Robinson V, Petticrew M, Kristjansson E, Moher D, McGowan J, et al.: **Is health equity considered in systematic reviews of the Cochrane Collaboration?** 11th Cochrane Colloquium, Barcelona, Spain; 2003.
- Ogilvie D, Petticrew M: **Smoking policies and health inequalities.** *Tobacco Control* 2004, **13**:129-31.
- Mackenbach J: **Tackling inequalities in health: the need for building a systematic evidence base.** *Journal of Epidemiology and Community Health* 2003, **57**:162.
- Gruen R, Bailie R, McDonald E, Weeramanthri T, Knight S: **The potential of systematic reviews to identify diversity and ineq-**

- uity in health care interventions. XI Cochrane Colloquium: Evidence, health care and culture; Barcelona; 2003.
30. Tsikata S, Robinson V, Petticrew M, Kristjansson B, Moher D, McGowan J, et al.: **Do Cochrane systematic reviews contain useful information about health equity?** XI Cochrane Colloquium: Evidence, health care, and culture; 2003.
 31. Ogilvie D, Hamilton V, Egan M, Petticrew M: **Systematic reviews of health effects of social interventions: I. Finding the evidence: how far should you go?** *J Epidemiol Community Health* 2005, **59**:804-8.
 32. Thomson H, Petticrew M, Morrison D: **Housing interventions and health – a systematic review.** *British Medical Journal* 2001, **323**:187-90.
 33. Thomson H, Petticrew M, Douglas M: **Health impact assessment of housing improvements: incorporating research evidence.** *Journal of Epidemiology and Community Health* 2003, **57**:11-6.
 34. Cooper H, Hedges L, editors: **The handbook of research synthesis.** New York: Russell Sage Foundation; 1994.
 35. Shadish W: **Meta-analysis and the exploration of causal mediation processes: A primer of examples, methods, and issues.** *Psychological Methods* 1996, **1**:47-65.
 36. Davey Smith G, Egger M, Phillips A: **Beyond the grand mean?** *British Medical Journal* 1997, **315**:1610-4.
 37. Thompson S: **Why and how sources of heterogeneity should be investigated.** *Systematic Reviews in Health Care: Meta-analysis in context* 2001:157-175.
 38. Alderson P, Green S, Higgins JPT, (eds): **Investigating heterogeneity. Cochrane Reviewers' Handbook 4.2.3 [updated November 2004]; Section 4.** In *The Cochrane Library Issue 2* Chichester, UK: John Wiley & Sons, Ltd; 2005.
 39. Oxman AD, Guyatt GH: **A consumers guide to subgroup analyses.** *Annals of Internal Medicine* 1992, **116**:78-84.
 40. Oxman AD, Guyatt G: **When to believe a subgroup analysis.** In *Users' Guide to the Medical Literature. A Manual for Evidence-Based Clinical Practice* Edited by: Guyatt G, Rennie D. Chicago: AMA Press; 2002:553-65.
 41. Brookes S, Whitley E, Peters T, Mulheran P, Egger M, Davey Smith G: **Subgroup analyses in randomised controlled trials: quantifying the risks of false positives and false negatives.** *Health Technology Assessment* 2001, **5**: [<http://www.hta.nhsweb.nhs.uk/htapubs.htm>].
 42. Dans A, McAlister F, Dans L, Richardson WS, Straus S, Guyatt G: **Applying the results to individual patients.** In *Users' Guides to the Medical Literature* Edited by: Guyatt G, Rennie D. Chicago: AMA Press; 2002:369-84.
 43. Macintyre S, Chalmers I, Horton R, Smith R: **Using evidence to inform health policy: case study.** *BMJ* 2001, **322**:222-5.
 44. Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.** .
 45. Panisset U: **A Review of WHO recommendations published in 2005.** .
 46. Kelly M, Swann C, Killoran A, Naidoo B, Barnett-Paige E, Morgan A: **Methodological Problems in Constructing the Evidence Base in Public Health.** Health Development Agency Public Health Evidence Steering Group Methodology Reference Group. Starter Paper. 9 August 2002. .

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation

Holger J Schünemann*¹, Atle Fretheim² and Andrew D Oxman²

Address: ¹INFORMA/CLARITY Research Group, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and
²Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Holger J Schünemann* - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@kunnskapssenteret.no; Andrew D Oxman - oxman@online.no
* Corresponding author

Published: 08 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:25 doi:10.1186/1478-4505-4-25

Accepted: 08 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/25>

© 2006 Schünemann et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the thirteenth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on applicability, transferability, and adaptation of guidelines.

Methods: We searched five databases for existing systematic reviews and relevant primary methodological research. We reviewed the titles of all citations and retrieved abstracts and full text articles if the citations appeared relevant to the topic. We checked the reference lists of articles relevant to the questions and used snowballing as a technique to obtain additional information. We used the definition "coming from, concerning or belonging to at least two or all nations" for the term international. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: We did not identify systematic reviews addressing the key questions. We found individual studies and projects published in the peer reviewed literature and on the Internet.

Should WHO develop international recommendations?

- Resources for developing high quality recommendations are limited. Internationally developed recommendations can facilitate access to and pooling of resources, reduce unnecessary duplication, and involve international scientists.
- Priority should be given to international health problems and problems that are important in low and middle-income countries, where these advantages are likely to be greatest.
- Factors that influence the transferability of recommendations across different settings should be considered systematically and flagged, including modifying factors, important variation in needs, values, costs and the availability of resources.

What should be done centrally and locally?

- The preparation of systematic reviews and evidence profiles should be coordinated centrally, in collaboration with organizations that produce systematic reviews. Centrally developed evidence profiles should be adaptable to specific local circumstances.
- Consideration should be given to models that involve central coordination with work being undertaken by centres located throughout the world.
- While needs, availability of resources, costs, the presence of modifying factors and values need to be assessed locally, support for undertaking these assessments may be needed to make guidelines applicable.
- WHO should provide local support for adapting and implementing recommendations by developing tools, building capacity, learning from international experience, and through international networks that support evidence-informed health policies, such as the Evidence-informed Policy Network (EVIPNet).

How should recommendations be adapted?

- WHO should provide detailed guidance for adaptation of international recommendations.
 - Local adaptation processes should be systematic and transparent, they should involve stakeholders, and they should report the key factors that influence decisions, including those flagged in international guidelines, and the reasons for any modifications that are made.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the thirteenth of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Adaptation involves modification according to different circumstances or environmental conditions [1]. In the context of guidelines, it relies on judgments of whether a guideline is applicable (i.e. relevant to a local setting in a specific setting) or transferable from one setting to another. A survey of managed care plans in the US found that they relied on national and other published guidelines as references for their own guidelines. However, most of the surveyed plans did not adopt published guidelines "as is" and adapted them for a variety of reasons [2]. The main reasons were lack of local clinical input, inappropriate consideration of resources, failure to apply to a specific population, too extensive recommendations, a high level of complexity in guidelines for users, and failure to include the most recent information in guidelines.

Article II of the World Health Organization (WHO) Constitution defines "setting, validating, monitoring and pursuing the proper implementation of norms and standards" as core functions of the WHO [3]. Accordingly, WHO issues guidelines (for example [4]) that are being used in many countries with the aim of improving the

quality of patient care and public health throughout the world [5]. In addition to WHO, an increasing number of organizations develop guidelines. Some of these organizations target international users. For this article we define the term "international" as "concerning or belonging to at least two or all nations". Among organizations that develop guidelines, WHO has the broadest mandate and spectrum of international consumers and stakeholders, given 192 countries are members of WHO and the scope of WHO's responsibilities.

Developing guidelines internationally poses challenges to ensure and monitor that WHO's guidelines are locally applicable or adaptable across different settings. The needs for adapting guidelines identified in the US survey referred to above are even greater for international guidelines. Organisations such as WHO that develop international guidelines need to consider variations in the contexts in which the guidelines will be applied, including differences in needs, values and the availability of resources. In this paper we addressed the following questions:

- Should WHO develop international recommendations?
- What should be done centrally and locally?
- How should recommendations be adapted?

Questions related to guideline implementation and dissemination are specifically addressed in another paper in this series [6].

What WHO is doing now

The Guidelines for WHO Guidelines (GWG) note that "Governments have as their main responsibility the health of the population, rather than the disease of the individual, and must consider other factors in addition to the traditional concern for maximizing the benefit to individual patients. WHO needs to assess the implications for population health of any recommendation as well. This requires explicit recognition that resources to provide health interventions (including diagnostic procedures, pharmaceuticals, surgical interventions and psychosocial techniques) are limited. This involves considering the cost-effectiveness of alternative interventions, the opportunity costs of investing in one intervention versus another, the affordability of the interventions, and the feasibility of applying a set of recommendations in different settings [7]."

The GWG also recognizes that "WHO takes a global perspective in addressing the needs of (192) member states. Differences in outcome will not only be due to transferring results from a research to a field setting, but also from the different cultural, economic, socio-demographic contexts present in the member-states," and that for "WHO guidelines, the traditional approach of reviewing and reporting evidence on efficacy and safety is certainly crucial but not sufficient. It can be regarded as the first step, but it is also necessary to examine the implications of applying each possible set of recommendations on a population basis. The initial body of evidence to be considered in WHO guidelines will be identical to that of traditional guidelines, but WHO guidelines will need to go further, to take the second step of spelling out the implications of adopting recommendations on costs and on population health. If done adequately, this will allow decision makers in different settings to take the third step of "localizing" the guidelines to their settings, and deciding where the trade-off between additional benefit and additional costs should be set. It will also be useful in determining what is acceptable for the end-users." However, one of the limitations of the GWG is that they have not been operationalized or implemented consistently. Only few WHO guideline processes have followed the GWG [8].

Indeed, the need for international development with local adaptation is expressed in WHO statements such as "The strategy recommends a prevention-oriented approach that emphasizes the need for countries to develop coherent, multi-sectoral national strategies with a long-term, sustainable perspective, to make the healthy choices the preferred alternatives at both the individual and community level. We welcome the commitment shown by Member States to the strategy and will be working closely with them to help them implement its recommendations [9]."

For example, the Global Strategy on Diet, Physical Activity and Health states that "the purpose of the Regional Consultations with Member States is for countries in each region to provide information on the extent of the problem associated with diet, physical activity and chronic disease, and appropriate prevention strategies for their particular countries. The consultation will focus on the discussion of national, regional and international interventions that will be effective within individual countries and that will take account of national, social, cultural and economic realities. Regional differences, common concerns, or international consensus, will be noted and serve as the basis of the development of the Global Strategy. This consultation process will build on past and current activities and programmes on the issue carried out by WHO Regional Offices and by Member States [10]."

WHO also provides funding and support for specific guideline adaptation efforts. Specific WHO guideline adaptation projects exist in the area of HIV that have been supported by international workshops organized by regional offices with country involvement [11,12]. Another example is a WHO-sponsored conference by the International Council of Ophthalmology on local adaptation of clinical practice guidelines in China [13]. However, while WHO is developing international guidelines through a variety of efforts, few WHO groups are using systematic and transparent processes that facilitate judgements regarding their applicability and transferability or provide guidance about how to adapt the guidelines [14].

What other organisations are doing?

We are not aware of published surveys that address what other organisations do to ensure appropriate adaptation of guidelines. The items that are part of the AGREE instrument [15] include the following three items most relevant for the assessment of guideline applicability:

- The potential organisational barriers in applying the recommendations should be discussed.
- The potential cost implications of applying the recommendations should be considered.
- The guideline should present key review criteria for monitoring and audit purposes.

The Conference on Guideline Standardization (COGS) checklist for reporting clinical practice guidelines suggests that guidelines should: "Describe the intended users of the guideline (e.g., provider types, patients) and the settings in which the guideline is intended to be used" [16]. The checklist does not include any specific recommendations related to supporting judgements about the applica-

bility or transferability of guidelines, or their local adaptation.

There is, however, a growing interest in considerations of how to adapt guidelines [17-22]. This interest is driven by several factors, including a desire to reduce unnecessary duplication of efforts across organisations, limited resources for many organisations, particularly in low and middle-income countries (LMIC), and concerns about the sustainability of programs that are well resourced.

In this background section we provide selected examples of organizations that have specifically provided information relevant to the key questions we posed although the WHO is unique in that its mandate includes the more complex task of providing international guidance.

SIGN

The Scottish Intercollegiate Guideline Network (SIGN) asks guideline panels to consider issues of applicability when guideline groups summarize their view of the total body of evidence [23]. The guidelines are graded to differentiate between those based on strong evidence and those based on weak evidence. This judgement is made on the basis of a transparent assessment of the design and quality of each study but also a judgement on the consistency, clinical relevance and external validity of the whole body of evidence. The aim is to produce a recommendation that is evidence-based, but which is relevant to the way in which health care is delivered in Scotland and is therefore implementable. The following specific points are included in the described considered judgments SIGN panels are asked to make:

- Generalisability of study findings
- Directness of application to the target population for the guideline.
- Clinical impact (i.e. the extent of the impact on the target patient population, and the resources needed to treat them.)
- Implementability (i.e. how practical it would be for the NHS in Scotland to implement the recommendation.)

SIGN guideline development groups are provided with a form in which to record the main points from their considered judgement [24]. Once they have considered these issues, the group is asked to summarise its view of the evidence and assign a level of evidence to it, before going on to derive a graded recommendation. During this process SIGN guideline developers are also able to downgrade a recommendation if they think the evidence is not generalisable, not directly applicable to the target population, or

for other reasons is perceived as being weaker than a simple evaluation of the methodology would suggest. In other areas, the appropriate action may be inclusion in the guideline of a commentary on the main economic issues that should be considered in relation to the subject of the guideline (for example [25]). Another option is the provision of basic information that will allow guideline users to work out the resource implications for their own service (for example [26]).

New Zealand Guideline Group

The process recommended by the New Zealand Guideline Group includes the following steps (figure 1) when adapting overseas evidence-based guidelines [27]:

- Appraise the guidelines (using AGREE instrument) for quality and process
- Analyse the content for scope and applicability
 - Same health settings, professional groups?
 - Same patients, consumers?
 - Same interventions?
 - Same outcomes?
- Look at the gaps in the issues covered in the overseas guidelines
 - Any clinical questions not covered?
 - Look at the sources of evidence
 - Is the search strategy available?
 - Are there any evidence tables?
- Are the evidence statements and recommendations referenced?
 - Re-run the search strategy to include the questions selected
 - To include literature at least one year prior to the date of publishing
 - Check if any large study would radically change the recommendations
- Implementation planning
 - Redesign the implementation plan to meet local circumstances

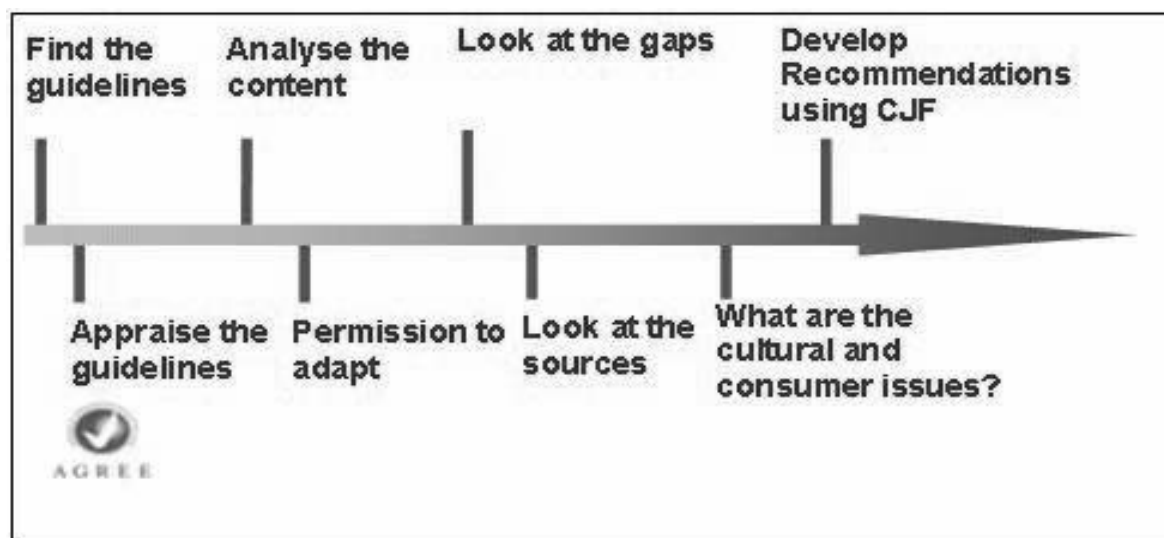


Figure 1
Adapting guidelines using the AGREE instrument (from [23]).

Guide to Community Preventive Services

The Guide to Community Preventive Services states the following in a discussion of its methods [28]: "The *Guide* should not be viewed as the sole source for informed public health decision making because local contextual information is also important. Many issues not addressed in the *Guide* will affect which interventions are implemented (e.g., resource availability, social justice, community participation, cultural appropriateness, local burden of diseases and risk factors, and political considerations). However, the *Guide* provides systematically collected and detailed information on several issues of importance to public health practitioners and decision makers; information which is difficult or inefficient to develop locally. *Guide* reviews and recommendations will be most useful in conjunction with a participatory community planning process that clarifies needs and goals and that considers the *Guide's* evidence reviews and recommendations in conjunction with additional applicable community specific information."

Methods

The methods used to prepare this review are described in the introduction to this series [29]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). For this review we analyzed existing guidelines for guidelines of national or interna-

tional organizations to identify processes that these organizations use to adapt guidelines locally beyond what was known for existing organizations as described in the background section. We also searched PubMed using "guideline" AND "adaptation OR applicability OR template OR transferability" (MESH headings/keywords) for studies and systematic reviews comparing different strategies to increase adaptation, acceptance and transferability (we identified 637 citations of which 203 citations were identified as systematic reviews using the clinical queries filter for systematic reviews). We reviewed the titles of all citations and retrieved abstracts and full text articles if the citations appeared relevant to the topic. We checked the reference lists of articles relevant to the questions and used snowballing as a technique to obtain additional information. We also searched the Cochrane Library and Google for articles and methods related to guideline adaptation ("guideline adaptation"). In addition, we searched databases maintained by the Agency for Healthcare Research and Quality (AHRQ, [30]) and the Guidelines International Network (GIN, [31]). The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Findings

We did not identify systematic reviews addressing the key questions. We found individual studies and projects pub-

lished in the peer reviewed literature and on the Internet that we will use to illustrate the responses to the key questions.

Should WHO develop international recommendations?

Threats of new emerging diseases (e.g. severe acute respiratory syndrome (SARS) and avian influenza A (H5N1) infection) as well as pandemics of chronic diseases such as obesity and heart disease have prompted international action and are clear examples of the existence of international health problems and the need for international recommendations. Given the international orientation of WHO and the advantages of large international organizations (e.g. accessing and pooling of resources, reducing unnecessary duplication, and involving international scientists), there is an important role for international recommendations [32]. International recommendations may be most helpful when variation in settings and local circumstances is less important. Therefore, consideration of need (prevalence, baseline risk or health status), setting (e.g. availability of resources) and modifying factors (factors that modify translation of recommendation into practice such as microbiological resistance patterns) can be key components that influence the strength of a recommendation and should be specified in recommendations formulated by the WHO [33].

An illustration of the need for adaptation is provided by Rhinehart and colleagues who attempted to implement a nosocomial infection control program based on the US Centers for Disease Control (CDC) guidelines in an urban Indonesian public hospital [34]. Adoption of unmodified CDC guidelines was impeded by modifying factors such as conditions of the physical plant, absence of an infection control infrastructure, limited sterilization capabilities, lack of clinical microbiologic laboratory support, and the expense of single use medical devices. After on-site evaluations, CDC guidelines were extensively modified so that they were appropriate for local conditions and culture [34]. After implementation, many physical changes had been accomplished, and handling of reusable and disposable medical devices had improved considerably although adoption of clinical practice policies was incomplete.

Global recommendations should apply to most settings yet allow for adaptation to local circumstances. The factors that influence recommendations should be laid out explicitly. If differences in context are likely to lead to different recommendations or decisions, these should be flagged [33,35]. Contextual issues that should be considered include modifying factors, need, values and resources. Table 1 provides a checklist of factors that influence the applicability or transferability of guidelines. It can be used during the guideline development process to

help ensure that these factors are considered systematically and transparently, and to clearly label factors that are important to consider in specific settings where the guideline will be applied or adapted.

What should be done centrally and locally?

The research evidence on what reflects the best distribution of responsibilities during the development of international guidelines is sparse. Global evidence (i.e. the best evidence from around the world) is the best starting point for judgements about effects and likely modifying factors. Synthesizing and making available this evidence should be coordinated centrally, although the actual work can be done anywhere. For example, the Agency for Health Care and Quality (AHRQ) funds Evidence-based Practice Centres throughout the US and Canada with the methodological competency to undertake systematic reviews, the UK National Institute for Health and Clinical Excellence (NICE) funds National Collaborating Centres responsible for guideline development (in areas such as acute care, cancer and chronic conditions), and the Cochrane Collaboration has 50 Collaborative Review Groups spread around the world that are responsible for preparing and updating systematic reviews (in particular areas such as breast cancer, infectious diseases, and tobacco addiction). WHO could adapt, commission or prepare systematic reviews that are required for guideline development in collaboration with organizations such as these that conduct systematic reviews and follow suggestions to make these reviews more useful for policymakers [36]. Lavis and colleagues suggest that donors and international agencies can encourage more informed public policymaking by supporting national and regional efforts to undertake systematic reviews and assess their local applicability, and by supporting regional or worldwide efforts to coordinate review and assessment processes [37].

Similarly, adaptable evidence profiles [33] based on new or existing high quality systematic reviews that include information on critical outcomes should also be prepared or coordinated centrally. Applying the criteria listed in table 1 and flagging important factors that influence the applicability of guidelines in evidence profiles could facilitate local adaptation and help groups to replace the flagged elements with locally appropriate information. Because resources for guidelines development are limited, particularly in LMIC, support for local adaptation of guidelines should also be supported centrally, by WHO headquarters or by regional offices. Given that WHO also has limited capacity, consideration should be given to doing this collaboratively with other organizations and to developing capacity; e.g. through the development of frameworks and tools, such as those being developed by the International Clinical Epidemiology Network (INCLIN) Knowledge Plus Program [20], GIN [31] and

Table 1: Checklist for identifying guidelines requiring adaptation

Factors influencing the applicability or transferability of guidelines across different settings	Response (positive answers increase the likelihood that recommendations should be flagged as requiring adaptation)
1. Is there important variation in need (prevalence, baseline risk or health status) that might lead to different decisions?	<input type="checkbox"/> Yes <input type="checkbox"/> Unclear <input type="checkbox"/> No
2. Is there important variation in the availability of resources that might lead to different decisions?	<input type="checkbox"/> Yes <input type="checkbox"/> Unclear <input type="checkbox"/> No
3. Is there important variation in costs (e.g. of drugs or human resources) that might lead to different decisions?	<input type="checkbox"/> Yes <input type="checkbox"/> Unclear <input type="checkbox"/> No
4. Is there important variation in the presence of factors that could modify the expected effects (e.g. resistance patterns of microbiological pathogens), which might lead to different decisions?	<input type="checkbox"/> Yes <input type="checkbox"/> Unclear <input type="checkbox"/> No
5. Is there important variation in the relative values of the main benefits and downsides that might lead to different decisions?	<input type="checkbox"/> Yes <input type="checkbox"/> Unclear <input type="checkbox"/> No

others; through training; through networks such as the Evidence-Informed Policy Network (EVIPNet) [38], and by learning from the experience of organizations around the world that are engaged in supporting evidence-informed health policies in specific settings [39].

Apparently successful examples of collaboration between central and country level groups exist. Wabitsch and colleagues described that during the adaptation of the WHO global HIV/AIDS guidelines [11] standard techniques involving consensus building were successfully employed to adapt these guidelines to local settings (Malawi and Barbados). The results showed that the process preserved the structure but involved significant modification to the processes of clinical care. Given the factors that influence the formulation of recommendations, the modification of individual recommendations confirms that an adaptation process was required.

How should recommendations be adapted?

In addition to supporting appropriate adaptation of its own guidelines, WHO should consider adapting guidelines developed by other organizations, given the potential value of WHO endorsement and savings, if high quality guidelines already exist. Detailed guidance on appropriate methods for adapting guidelines would help WHO guideline groups to adapt existing guidelines, when this is appropriate. Adaptation of recommendations is required because several judgments influence recommendations. Therefore, recommendations dealing with identical questions may differ between developers despite reliance on the same evidence. Implementation, which follows the process of adaptation, is topic of another paper in this series and tools for the evaluation of implementation of guidelines have been developed [6,40].

Decisions during local adaptation processes should be transparent and follow procedures that are similar to those used in developing the guidelines, including reporting the key factors that influence any modifications. Two fairly similar approaches have recently appeared that produced frameworks for identifying candidate guidelines for local adaptation. The Practice Guideline Evaluation and Adaptation Cycle (PGEAC) is a 10 step approach (figure 2) [17,41,42]. Graham and colleagues describe three alternatives in the PGEAC approach: (a) adopt one guideline with all its recommendations; (b) adopt one guideline, endorsing some of its recommendations but not endorsing recommendations that lack strong evidence or cannot be implemented or adapted locally; or (c) take the best recommendations from each of the guidelines and adapt them for inclusion of the new guideline [17]. If recommendations are modified the rationale for changes should be explicitly stated in the resulting local guideline document.

The other approach has been developed by the international working group ADAPTE [18,19,22] and partly overlaps with the PGEAC approach. Recently the groups developing both adaptation approaches merged with the purpose of developing a generic manual on guideline adaptation. The manual will undergo pilot testing. The group is calling itself the ADAPTE group <http://www.adapt.org>. Whatever adaptation process is chosen the process should be made explicit, undergo review by peers, and involve consumers, policymakers and other stakeholders who may provide input about policy guidance.

Balance sheets or evidence profiles are designed to assist decision-makers regarding outcomes in their practice set-

PRACTICE GUIDELINES EVALUATION AND ADAPTATION CYCLE

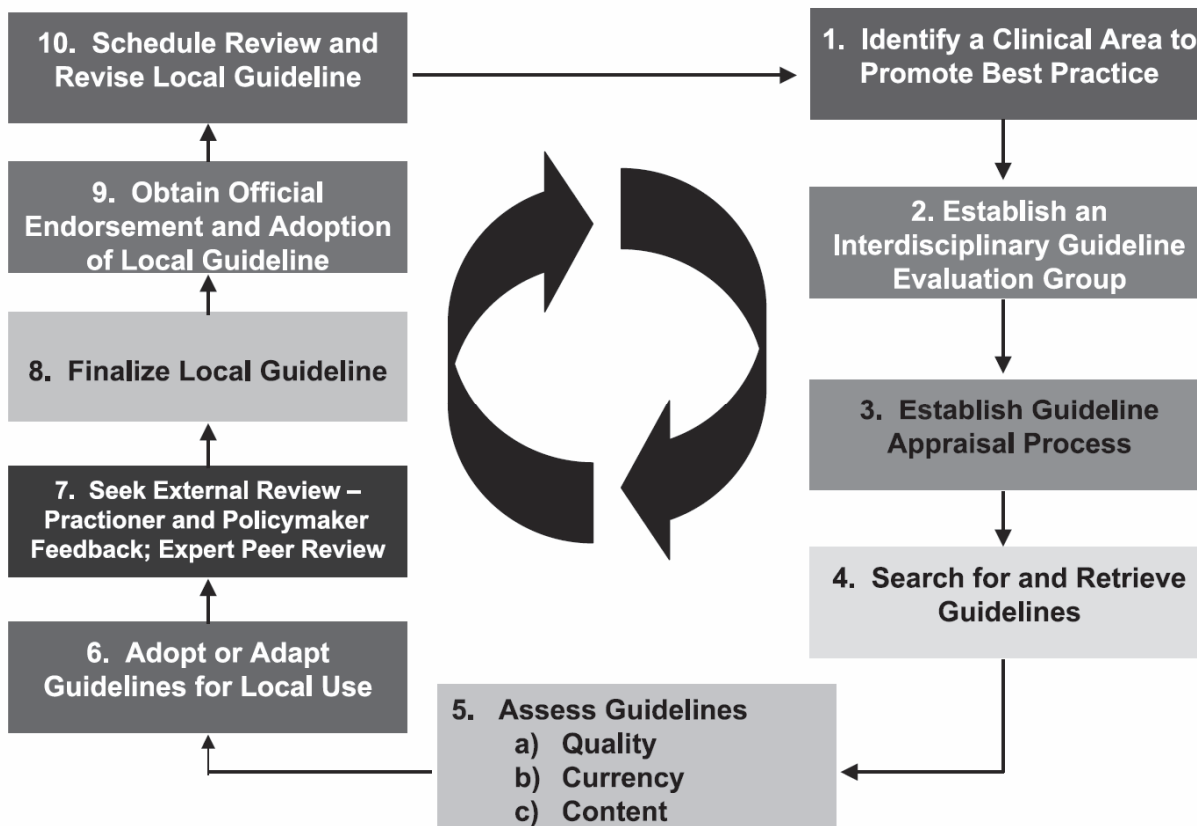


Figure 2
The Practice Guideline Evaluation and Adaptation Cycle (PGEAC) (from [17, 41]).

ting [43,44]. For guideline adaptation they should include data for the specific setting to which they are adapted (for all the considerations in table 1). During adaptation consumer involvement (i.e. to integrate their values and perspectives) and involvement of other stakeholders should be emphasized [35,45].

Discussion

Given WHO's mandate; limited resources that are available to develop high quality guidelines that are informed by the best available evidence, particularly in LMIC; and the potential to reduce unnecessary duplication, WHO should continue to develop international guidelines. However, these guidelines will often require adaptation and tailoring to local contexts and WHO should, so far as possible, provide support to help ensure that international guidelines are adapted appropriately to local circumstances. To do this WHO must ensure that it

systematically considers needs for local adaptation when developing guidelines; and that it has sufficient capacity to support both developing and supporting the adaptation of high priority guidelines. To do this as effectively and efficiently as possible WHO should collaborate with other organizations both in developing guidelines and, importantly, in developing capacity in LMIC and supporting appropriate adaptation in countries that lack resources. We provide other recommendations about how WHO can improve the implementation of organizational changes to guideline development in other articles in this series [6,46-50].

Further work

WHO has ample experience in adapting guidelines but this effort should be coordinated and disseminated among WHO guideline groups. Through coordinating guideline development within WHO and collaborating

with other organizations, WHO could capitalise on this experience, improve the quality of its guidelines, and help to ensure that its guidelines are appropriately adapted and result in appropriate actions and health improvements. Similar suggestions have been made previously by investigators involved with WHO guideline projects. For example, "Countries should discuss and find ways of collaboration and formation of linkages and support with National HIV/AIDS program in order to enhance the implementation of IMCI algorithm which includes HIV/AIDS" [11].

A systematic review of studies evaluating methods for adapting guidelines is unlikely to retrieve high quality evidence given the paucity of research in this area. Further development and evaluation of frameworks and tools to support the appropriate adaptation of guidelines is needed. Given the limited capacity for this in many countries, comparisons of simpler processes that require fewer resources should be compared with more rigorous processes to determine the most efficient methods for ensuring that guidelines support well-informed decisions and actions appropriate for the specific contexts in which they are taken.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

HJS prepared the first draft of this review. AF and ADO contributed to drafting and revising it. All authors read and approved the final manuscript.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series. The authors thank Mr. John Fox for administrative assistance and help with organizing the literature review.

References

1. www.wordreference.com: **Word reference dictionary**. [<http://www.wordreference.com/definition>].
2. **United States General Accounting Office. Report to the Chairman, Subcommittee on Health, Committee on Ways and Means, House of Representatives. Practice Guidelines. Managed Care Plans Customize Guidelines to Meet Local Interests**. 1996, **GAO/HEHS-96-95**.
3. **World Health Organization. Constitution of the World Health Organization**. *Off Rec Wld Hlth Org*, 2, 100 1948.
4. **World Health Organization. Prevention of Recurrent Heart Attacks and Strokes in Low and Middle Income Populations: Evidence-based Recommendations for Policy Makers and Health Professionals**. Geneva , World Health Organization; 2003.
5. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs**. *Int J Qual Health Care* 2003, **15(1)**:31-45.
6. Fretheim A Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 15. Disseminating and implementing guidelines**. *Health Res Policy Syst* 2006.
7. **Global Programme on Evidence for Health Policy. Guidelines for WHO Guidelines**. EIP/GPE/EQC/2003.1. Geneva , World Health Organization; 2003.
8. Schünemann HJ, Hill S, Kakad M, Bellamy R, Uyeki T, Hayden F, Yazdanpanah Y, Beigel J, Chotpitayusunondh T, Vist G, et al.: **WHO Rapid Advice Guidelines for the pharmacological management of sporadic human infection with avian influenza A (H5N1) virus**. *The Lancet Infectious Disease* 2007, **7**:21-31.
9. **World Health Organization (WHO). Global Strategy on Diet, Physical Activity and Health**. In *Bulletin of the WHO WHA57/A57_R17* Geneva , World Health Organization; 2002.
10. **World Health Organization. World Health Organization Process for a Global Strategy on Diet, Physical Activity and Health. WHO/NMH/EXR.02.2 Rev.1**. Geneva , World Health Organization; 2003.
11. Wabitsch R, Margolis CZ, Malkin JE, Abu-Saad K, Waksman J: **Evaluating the process of tailoring the global HIV/AIDS practice guidelines for use in developing countries**. *Int J Qual Health Care* 1998, **10(2)**:147-154.
12. **Report on the Workshop on Adaptation of IMCI Guidelines to Include HIV AIDS. Harare 18 to 23 June 2001** 2001 [http://www.who.int/child-adolescent-health/New_Publications/HIV_report_HIV_Harare.htm].
13. **ICO International Clinical Guidelines: Report on COS Preferred Practice Pattern Guidelines Development Workshop Beijing, China, June 24-26, 2005** [<http://www.icoph.org/guide/chinarp.html>]
14. **Personal communication with WHO Evidence in Policy - no date**.
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(1)**:18-23.
16. 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.
17. Graham ID, Harrison MB, Brouwers M, Davies BL, Dunn S: **Facilitating the use of evidence in practice: evaluating and adapting clinical practice guidelines for local use by health care organizations**. *J Obstet Gynecol Neonatal Nurs* 2002, **31(5)**:599-611.
18. Fervers B, Graham ID: **Guideline adaptation - the way forward**. In *Guidelines International Network Conference 2005* Lyon; 2005.
19. Burgers JS, Cook R: **Adaptation of Guidelines**. *Guidelines International Network: 2005* 2005 [http://www.g-i-n.net/download/files/Rob_Cook_Adaptation_of_guidelines.pdf]. Lyon
20. **INCLEN Knowledge 'Plus' Program (K+). The INCLEN Trust**. [<http://www.inclentrust.org>]
21. **Adapting existing HTAs from one country to other. European Network for Health Technology Assessment - EUnetHTA** [http://www.sst.dk/upload/planlaegning_og_behandling/centv/eunethta/wp5description.pdf]
22. Fervers B Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L, Coulombe M, Poirier M, Burnand B: **Adaptation of clinical guidelines: literature review and proposition for a framework and procedure**. *Int J Qual Health Care* 2006, **13**:167-176.
23. Yuen KY: **Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus**. *Lancet* 1998, **351**:467-471.
24. **Scottish Intercollegiate Guideline Network. SIGN 50: A guideline developers' handbook. Section 5**. (accessed on

- March 17, 2006) [<http://www.sign.ac.uk/guidelines/fulltext/50/section5.html>].
25. **Scottish Intercollegiate Guideline Network. Cardiac Rehabilitation.** In *SIGN Publication Volume No. 57*. Edinburgh, Scottish Intercollegiate Guideline Network; 2002.
 26. **Scottish Intercollegiate Guideline Network.** In *Management of osteoporosis Volume SIGN Publication No. 71*. Edinburgh, Scottish Intercollegiate Guideline Network; 2002.
 27. **New Zealand Guidelines Group. Notes on the Adaptation / Synthesis of Guidelines (21-Jan-04)** [<http://www.nzgg.org.nz/>]
 28. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Agüero L, Truman BI, Hopkins DP, Mullen PD, Thompson RS, Woolf SH, Carandek-Kulis VG, Anderson L, Hinman AR, McQueen DV, Teutsch SM, Harris JR: **Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services.** *Am J Prev Med* 2000, **18(1 Suppl)**:35-43.
 29. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: Introduction.** *Health Res Policy Syst* 2006, **4**:12.
 30. **Agency for Healthcare Research and Quality** [<http://www.guidelines.gov>]
 31. **Guidelines International Network** [<http://www.g-i-n.net>]
 32. Gaebel W, Weinmann S, Sartorius N, Rutz W, McIntyre JS: **Schizophrenia practice guidelines: international survey and comparison.** *Br J Psychiatry* 2005, **187**:248-255.
 33. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006, **4**:20.
 34. Rhinehart E, Goldmann DA, O'Rourke EJ: **Adaptation of the Centers for Disease Control guidelines for the prevention of nosocomial infection in a pediatric intensive care unit in Jakarta, Indonesia.** *Am J Med* 1991, **91(3B)**:213S-220S.
 35. Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 10. Integrating values and consumer involvement.** *Health Res Policy Syst* 2006, **4**:22.
 36. Gruen RL, Morris PS, McDonald EL, Bailie RS: **Making systematic reviews more useful for policy-makers.** *Bull World Health Organ* 2005, **83(6)**:480.
 37. Lavis JN, Posada FB, Haines A, Osei E: **Use of research to inform public policymaking.** *Lancet* 2004, **364(9445)**:1615-1621.
 38. Hamid M Bustamante-Manaog T, Dung T, Akkhavong K, Fu H, Ma Y: **EVIPNet: translating the spirit of Mexico.** *Lancet* 2006, **366**:1758-1760.
 39. 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.
 40. 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 Med Inform Decis Mak* 2005, **5**:23.
 41. Graham ID, Harrison MB, Lorimer K, Piercianowski T, Friedberg E, Buchanan M, Harris C: **Adapting national and international leg ulcer practice guidelines for local use: the Ontario Leg Ulcer Community Care Protocol.** *Adv Skin Wound Care* 2005, **18(6)**:307-318.
 42. Graham ID, Harrison MB, Brouwers M: **Evaluating and Adapting Practice Guidelines for Local Use: A Conceptual Framework.** In *Clinical Governance in Practice* Edited by: Pickering S, Thompson J. London, Harcourt Publishers; 2003:213-229.
 43. Braddick M, Stuart M, Hrachovec J: **The use of balance sheets in developing clinical guidelines.** *J Am Board Fam Pract* 1999, **12(1)**:48-54.
 44. 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: **Grading quality of evidence and strength of recommendations.** *Bmj* 2004, **328(7454)**:1490.
 45. Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 3. Group composition.** *Health Res Policy Syst* 2006, **4**:15.
 46. Oxman AD, Fretheim A, Schünemann AD: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006, **4**:19.
 47. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 14. Reporting guidelines.** *Health Res Policy Syst* 2006, **4**:26.
 48. Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 16. Evaluation.** *Health Res Policy Syst* 2006, **4**:28.
 49. Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 1. Review of guidelines for guidelines.** *Health Res Policy Syst* 2006, **4**:13.
 50. Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 9. Grading evidence and recommendations.** *Health Res Policy Syst* 2006, **4**:21.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: I4. Reporting guidelines

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim³

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway, ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy and ³Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 08 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, 4:26 doi:10.1186/1478-4505-4-26

Accepted: 08 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/26>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 14th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on reporting guidelines and recommendations.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: There is little empirical evidence that addresses these questions. Our answers are based on logical arguments and standards put forward by other groups.

What standard types of recommendations or reports should WHO use?

- WHO should develop standard formats for reporting recommendations to facilitate recognition and use by decision makers for whom the recommendations are intended, and to ensure that all the information needed to judge the quality of a guideline, determine its applicability and, if needed, adapt it, is reported.
- WHO should develop standard formats for full systematically developed guidelines that are sponsored by WHO, rapid assessments, and guidelines that are endorsed by WHO.
- All three formats should include the same information as full guidelines, indicating explicitly what the group preparing the guideline did not do, as well as the methods that were used.
- These formats should be used across clinical, public health and health systems recommendations.

How should recommendations be formulated and reported?

- Reports should be structured, using headings that correspond to those suggested by the Conference on Guideline Standardization or similar headings.
- The quality of evidence and strength of recommendations should be reported explicitly using a standard approach.
- The way in which recommendations are formulated should be adapted to the specific characteristics of a specific guideline.
- Urgent attention should be given to developing a template that provides decision makers with the relevant global evidence that is needed to inform a decision and offers practical methods for incorporating the context specific evidence and judgements that are needed.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 14th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Guidelines are formal advisory statements that should be robust enough to meet the unique circumstances and constraints of the specific situation to which they are being applied [1]. The basic nature and intent of guidelines have been variously labelled as guidance, guides, guiding principles, recommendations, policies, protocols, best practice, algorithms, consensus statements, expert committee recommendations, integrated care pathways, manuals, tool kits, handbooks, model lists, technical updates and principles [1,2]. Whatever they are called, rigorously developed guidelines, can translate complicated research findings into actionable recommendations. They are an important step in moving from research to action and ensuring that the best available research evidence informs decisions and actions [3]. However, for users of guidelines to be able to apply criteria to assess whether guidelines have been rigorously developed and are likely to be valid and applicable [3-6], the information needed to make these judgements must be reported [7]. Unfortunately, critical information is often absent from published guidelines [5,7,8].

In this paper we address the following questions:

- What standard types of recommendations or reports should WHO use?
- How should recommendations be formulated and reported?

We address questions about reporting systematic reviews [9] and dissemination [10] in other papers in this series.

What WHO is doing now

Although the Guidelines for WHO Guidelines recommends "that a uniform, readily-recognizable printing format be developed for WHO guidelines," there are, as yet, no standard formats for WHO policies, recommendations or guidelines. A survey of WHO guidelines published in 2005 found that WHO publishes a large number of recommendations of many different types, in many different formats [2], and a review of WHO documents did not find any standards for reporting WHO recommendations [11].

What other organisations are doing

In a recent survey of organisations that produce clinical practice guidelines, all 31 organisations that responded (response rate 86%), and 46 of 57 (81%) of units that support the use of research evidence by governments in developing health policy, reported producing full versions of guidelines with references and notes [12]. Several organisations use different formats for different types of recommendations, and a majority produce different versions of guidelines, such as executive summaries, summaries of take-home messages, separate versions for different target users, and tools for application (e.g., algorithms or flow charts). Many guideline producers have standard formats that they use and some organisations, such as the U.S. National Guidelines Clearing House, have developed standard formats for reporting guidelines produced by other organisations [13].

The UK National Center for Health and Clinical Excellence (NICE) states that recommendations should be clear and concise, but should contain sufficient information that they can be understood without reference to other supporting material [14]. This is particularly important where recommendations are published in isolation from the background details in the full guideline. Any terminology included in the recommendations therefore needs to be clearly defined and unambiguous.

Methods

The methods used to prepare this review are described in the introduction to this series [15]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register [16], the US National Guideline Clearinghouse [17], and the Guidelines International Network [18]) for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review we searched PubMed using [the MeSH terms 'Documentation/standards' and 'Practice Guidelines/standards'] and related articles; the Cochrane Methodology Register using [the key words 'CMR: Review methodology – applicability & recommendations' and the text words (format or reporting)] and ['Levels of evidence and strength of recommendations']; the National Guidelines Clearinghouse annotated bibliography using the terms format, reporting and structure; and checked the reference lists of retrieved articles.

Findings

What standard types of recommendations or reports should WHO use?

Given the wide variety of different types of recommendations that are made by WHO, there is likely to be a need for several standard types of recommendations. Systematically developed clinical practice guidelines can take 18 months or more and as many as 15 meetings [19]. Systematically developed public health guidelines also require substantial resources and time [20]. Given the time and resources required to produce guidelines, many organisations, particularly HTA organisations, have developed rapid assessment processes [21-25]. There is variation in the scope, methods, time to complete assessments, and the formats used to report rapid assessments. Another type of recommendation or guideline that is receiving increasing attention, also because of the resources and time required to develop guidelines systematically, are guidelines developed by other organisations that have been adapted or endorsed [26-28]. Another approach being taken in several countries is to create databases or clearinghouses of clinical practice guidelines with the aim of facilitating their evaluation and adaptation for local use by health care organizations [29].

Systematically developed clinical recommendations, public health recommendations, and health systems recommendations all require similar processes to ensure their quality. Decision makers also require similar types of information to be able to critically appraise whether guidelines have been rigorously developed and are likely to be valid [3,4,7,20,30,31].

How should recommendations be formulated and reported?

The Conference on Guideline Standardization (COGS) used a two-stage modified Delphi process to develop standards for reporting clinical practice guidelines [7]. Representatives of 22 organisations active in guideline development reviewed the proposed items and commented favourably. The items were consolidated into 18 topics (Table 1) to create the COGS checklist, which provides a framework to support comprehensive documentation of guidelines. While it is possible that some guideline developers may not include content for every item, it is suggested that they should address explicitly whether the guideline development team considered that item.

While many organisations have their own standard formats for reporting guidelines, this is the only consensus standard for reporting guidelines across organisations. We have, however, summarised the key items included in guidelines for guidelines, which provides the basis for a similar, but more comprehensive checklist for conducting or reporting guidelines [32]. In addition, there are a

number of instruments for evaluating clinical practice guidelines that can also be used as checklists for reporting [1,4-6,33]. The content used in the National Guideline Clearing house also represents a standard for reporting imposed on organisations that want their guidelines included in that database [13]. It includes 52 items under the following headings: scope, methodology – including rating scheme and cost analysis, recommendations, evidence supporting the recommendations, benefits/harms of implementing the recommendations, contraindications, qualifying statements, implementation of the guideline, Institute of Medicine (IOM) national health-care quality report categories, identifying information and availability, and disclaimer; in addition to indexing attributes.

Similarly, some journals have standard formats for reporting clinical practice guidelines, including structured abstracts with the following headings [34]:

Objective

a succinct statement of the objective of the guideline, including the targeted health problem, the targeted patients and providers, and the main reason for developing recommendations concerning this problem for this population.

Options

principal practice options that were considered in formulating the guideline.

Outcomes

significant health and economic outcomes identified as potential consequences of the practice options.

Evidence

methods used to gather, select, and synthesize evidence, and the date of the most recent evidence obtained.

Values

persons and methods used to assign values (relative importance) to potential outcomes of alternative practice options.

Benefits, harms, and costs

the type and magnitude of the main benefits, harms, and costs that are expected to result from guideline implementation.

Recommendations

a brief and specific list of key recommendations.

Table 1: The COGS checklist for reporting clinical practice guidelines (from Shiffman et al. [7])*

Topic	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 is 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 sources/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, an 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.

*COGS = Conference on Guideline Standardization.

Validation

the results of any external review, comparison with guidelines developed by other groups, or clinical testing of guideline use.

Sponsors

key persons or groups that developed, funded, or endorsed the guideline.

While many organisations have standards for how recommendations are formulated, we are not aware of any consensus standards for how recommendations should be formulated. Most guidelines development groups now grade the quality of evidence and the strength of recommendations, but a variety of different grading systems are used [12,35].

Shekelle and colleagues evaluated the effect of different levels of specificity of recommendations on clinicians test ordering behaviour using clinical vignettes [36]. They found that clinicians receiving the non-specific recommendations ordered fewer indicated tests for appropriate clinical vignettes than did physicians receiving specific recommendations. Furthermore, compared to physicians receiving non-specific guidelines, physicians receiving specific guidelines ordered significantly more appropriate tests for corresponding vignettes and significantly fewer tests for inappropriate vignettes. The authors concluded that the clarity and clinical applicability of a guideline might be important attributes that contribute to the effects of practice guidelines. We did not find any other comparisons of different ways of formulating recommendations, and it is likely that the way in which recommen-

dations are formulated may need to be adapted to the specific characteristics of a guideline. However, there is a consensus that recommended actions should be stated precisely [7].

Both checklists for evaluating guidelines and for reporting guidelines include items that may be dependent on the specific setting in which a guideline is developed or used. There are a number of international organisations, including WHO, that develop guidelines that are intended to be used in a variety of settings around the world. We did not find any published papers that addressed methods for taking into account setting specific factors in international guidelines, although several groups are working on methods for adapting guidelines developed in one setting for use in another [27,28,37-39].

Discussion

While the content of WHO guidelines, recommendations and policies will vary, depending on the topic, it would be desirable to have standard formats across different topics to facilitate recognition and use by decision makers and ensure that all the information needed to judge the quality of a guideline, determine its applicability and, if needed, adapt it is reported. There is likely a need for at least three standard formats: full systematically developed guidelines that are sponsored by WHO, rapid assessments, and guidelines that are endorsed by WHO. Standards such as those advocated by COGS should form the basis for developing a uniform format for full guidelines developed by WHO. Although the COGS standards were developed for clinical practice guidelines, the same considerations are relevant to public health and health systems recommendations. Further consideration is needed regarding the inclusion of additional items that need to be considered in WHO guidelines, such as applicability to different settings, equity, and scaling up. In particular, work is needed to develop a template for decision-making frameworks when different conditions are likely to lead to different decisions in different settings [40,41]. In addition, different versions of guidelines should be developed for different target audiences, including a structured executive summary and key messages [31,34].

A different format for rapid assessments would help to distinguish these from full guidelines and could be designed to reduce the work and time necessary to complete a report. Rapid assessments should, nonetheless, include the same information as full guidelines, indicating explicitly what the group preparing the guideline did not do, as well as the methods that were used.

WHO has limited resources and capacity for developing guidelines. At the same time, low and middle-income (LMIC) countries also have limited resources and capac-

ity, and a core function of WHO is to provide its member states, particularly LMIC, with technical advice that is informed by the best available research evidence. Through collaborating with other organisations and establishing standards for reporting, and possibly endorsing guidelines developed by other organisations, WHO may be able to expand its potential for supporting access to guidelines that are appropriate for LMIC or can easily be adapted to conditions in LMIC.

Given WHO's role as the world's leading public health agency and its mandate to provide evidence-informed guidance, it is of major importance for WHO to address and develop methods for formulating and reporting recommendations that are developed internationally, but need to be adapted and implemented in specific settings.

Further work

As part of the methodological work that is needed to address the challenges of developing international guidelines, specific attention should be given to the development of standard templates that provide decision makers with the relevant global evidence that is needed to inform a decision and offers practical methods for incorporating the context specific evidence and judgements that are needed [40,41].

In addition to ensuring that standard formats are used for WHO guidelines to ensure complete reporting, attention should be paid to ensuring that the format that is used, and derivative versions, are understandable and useful to the intended target audiences.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Global Programme on Evidence for Health Policy: **Guidelines for WHO Guidelines**. Geneva: World Health Organisation, 2003 (EIP/GPE/EQC/2003.1).
- Panisett U: **A review of WHO recommendations published in 2005**.
- Guyatt G, Hayward R, Richardson WS, Green L, Wilson M, Sinclair J, et al.: **Moving from evidence to action**. In *Users' Guides to the Medical Literature* Edited by: Guyatt G, Rennie D. Chicago: AMA Press; 2002:175-99.
- Institute of Medicine: **Guidelines for Clinical Practice: From Development to Use**. Washington, DC: National Academy Press; 1992.
- 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.
- The 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.
- 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**:493-8.
- 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.
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence**. *Health Res Policy Syst* 2006.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 15. Disseminating and implementing guidelines**. *Health Res Policy Syst* 2006.
- Oxman AD, Fretheim A, Schünemann HJ: **A review of WHO documents that guide the development, dissemination and implementation of recommendations by WHO**.
- 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.
- Guideline Summary Sheet. National Guideline Clearinghouse** [<http://www.guideline.gov/about/CompleteSummaryDescription.aspx>]
- What are the key components of a template for recommendations could be the question here?** *National Institutes for Health and Clinical Excellence Handbook* [http://www.nice.org.uk/pdf/GDM_Allchapters_0305.pdf].
- Oxman AD, Schünemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: introduction**. *Health Res Policy Syst* 2006.
- Cochrane Methodology Register** [http://www.cochrane.org/access_data/cmr/accessDB_cmr.asp]
- National Guideline Clearinghouse** [<http://www.guidelines.gov>]
- Guidelines International Network (GIN)** [<http://www.gin.net>]
- Raine R, Sanderson C, Black N: **Developing clinical practice guidelines: a challenge to current methods**. *BMJ* 2005, **331**:631-3.
- Briss PA, Zaza S, Pappaioanou M, et al.: **Developing an evidence-based guide to community preventive services-methods**. *Am J Prev Med* 2000, **18**(1S):35-43.
- Stevens A, Colin-Jones D, Gabbay J: **'Quick and clean': authoritative health technology assessment for local health care contracting**. *Health Trends* 1995, **27**:37-42.
- Hailey D, Corabian P, Harstall C, Schneider W: **The use and impact of rapid health technology assessments**. *Int J Technol Assess Health Care* 2000, **16**:651-6.
- Perleth M, Luhmann D, Givis B, Droste S: **Rapid Assessments – quick evaluation of medical technology**. *Gesundheitswesen* 2001, **63**(Suppl 1):S79-84.
- Canadian Coordinating Office for Health Technology Assessment: **Rapid response services in health technology assessment: information from INAHTA members and additional information**. Ottawa: CCOHTA; 2004.
- Milne R, Clegg A, Stevens A: **HTA responses and the classic HTA report**. *J Public Health Med* 2003, **25**:102-6.
- Schünemann H, Fretheim A, Oxman AD: **Adaptation, applicability, transferability**.
- Graham ID, Harrison MB, Brouwers M, Davies BL, Dunn S: **Facilitating the use of evidence in practice: evaluating and adapting clinical practice guidelines for local use by health care organizations**. *J Obstet Gynecol Neonatal Nurs* 2002, **31**:599-611.
- Graham ID, Harrison MB, Lorimer K, Piercianowski T, Friedberg E, Buchanan M, Harris C: **Adapting national and international leg ulcer practice guidelines for local use: the Ontario Leg Ulcer Community Care Protocol**. *Adv Skin Wound Care* 2005, **18**:307-18.
- Guideline Resources. National Guideline Clearinghouse** [http://www.guideline.gov/resources/guideline_resources.aspx]
- Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EAF, Millsom ME, et al.: **An approach to the development of practice guidelines for community health interventions**. *Can J Public Health* 1994, **85**(Suppl 1):S8-13.
- Lavis JN, Davies HTO, Oxman AD, Denis JL, Golden-Biddle K, Ferlie E: **Towards systematic reviews that inform healthcare management and policymaking**. *Journal of Health Services Research and Policy* 2005, **10**(Suppl 1):35-48.
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: I. Guidelines for guidelines**. *Health Res Policy Syst* 2006.
- Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM: **A comparison of clinical practice guideline appraisal instruments**. *Int J Technol Assess Health Care* 2000, **16**:1024-38.
- Hayward RS, Wilson MC, Tunis SR, Bass EB, Rubin HR, Haynes RB: **More informative abstracts of articles describing clinical practice guidelines**. *Ann Intern Med* 1993, **118**:731-7.
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 9. Grading evidence and recommendations**. *Health Res Policy Syst* 2006.
- Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M: **Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making**. *Health Serv Res* 2000, **34**:1429-48.
- INCLIN Knowledge 'Plus' Program (K+). The INCLIN Trust** [<http://www.inclintrust.org/index.php?option=content&task=view&id=182&Itemid=223#KPP>]
- Adapting existing HTAs from one country to other. European Network for Health Technology Assessment – EUnetHTA** [http://www.sst.dk/upload/planlaegning_og_behandling/cemtv/eunetha/wp5description.pdf]
- Fervers B, Graham I: **Guideline Guideline adaptation: the way forward?** *The Guidelines International Network* [http://www.g-i-n.net/download/files/Plenary_2_Fervers_Graham.pdf].
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include**. *Health Res Policy Syst* 2006.
- Schünemann HJ, Fretheim A, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 13. Adaptation, applicability and transferability**. *Health Res Policy Syst* 2006.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines

Atle Fretheim*¹, Holger J Schünemann² and Andrew D Oxman¹

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy

Email: Atle Fretheim* - atle.fretheim@nokc.no; Holger J Schünemann - hjs@buffalo.edu; Andrew D Oxman - oxman@online.no

* Corresponding author

Published: 08 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, **4**:27 doi:10.1186/1478-4505-4-27

Accepted: 08 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/27>

© 2006 Fretheim et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 15th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: In this review we address strategies for the implementation of recommendations in health care.

Methods: We examined overviews of systematic reviews of interventions to improve health care delivery and health care systems prepared by the Cochrane Effective Practice and Organisation of Care (EPOC) group. We also conducted searches using PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: What should WHO do to disseminate and facilitate the uptake of recommendations?

- WHO should choose strategies to implement their guidelines from among those which have been evaluated positively in the published literature on implementation research
- Because the evidence base is weak and modest to moderate effects, at best, can be anticipated, WHO should promote rigorous evaluations of implementation strategies.

What should be done at headquarters, by regional offices and in countries?

- Adaptation and implementation of WHO guidelines should be done locally, at the national or sub-national level.
- WHO headquarters and regional offices should support the development and evaluation of implementation strategies by local authorities.

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the 15th of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Developing recommendations makes little sense if they are not used. Thus, effective strategies to promote the appropriate use of recommendations by decision-makers (clinicians, public health officers, policymakers) are important.

In this paper we address the following questions:

- What should WHO do to disseminate and facilitate the uptake of recommendations?
- What should be done at headquarters, by regional offices and in countries?

Questions related to adaptation and evaluation of guidelines are addressed in other papers in this series [1,2].

What is WHO doing now?

There is no general WHO strategy for guideline implementation. The methods that are used vary from department to department, and may vary from case to case within departments. Field testing and rollout strategies that are used to promote the uptake of recommendations are often not informed by the findings of implementation research, and it is generally difficult to estimate the impact of the implementation strategies that are used, since evaluations are rarely rigorous, if they are done at all.

There are, however, examples of rigorous evaluations, such as implementation of the Integrated Management for Childhood Illnesses (IMCI) guideline, which has been evaluated in a randomised controlled trial [3]. WHO has also in some cases reviewed the relevant evidence-base, for example for strategies to improve the use of drugs in developing countries [4].

Although recommendations inevitably need to be adapted and implemented at country level, WHO headquarters and regional offices can support these activities [2].

What are other organisations doing?

In an international survey of organisations that develop guidelines or health technology assessments, almost half of the 95 respondents reported using provider-mediated

interventions as part of their strategy for implementing of guidelines [5]. Examples of this were conducting audits and hosting work-shops for practitioners. However, many respondents did not provide specific examples.

In a smaller international survey of prominent guideline developers, nearly all of the 18 organisations used educational materials and conferences as part of their implementation strategies [6]. Other common approaches were audit and feed-back, use of local opinion leaders, and organizational interventions (e.g. financial incentives or disincentives).

Methods

The methods used to prepare this review are described in the introduction to this series [7]. The key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). The Cochrane Effective Practice and Organisation of Care (EPOC) group undertakes systematic reviews of interventions to improve health care delivery and health care systems. EPOC has undertaken periodic overviews of systematic reviews to assess and summarise the evidence available from existing Cochrane and non Cochrane reviews [8-10]. The research findings reported here are drawn from these overviews and an update of those overviews that is underway. In addition, we searched PubMed and three databases of methodological literature (within the databases of The Cochrane Library, the US National Guideline Clearinghouse [11] and the Guidelines International Network [12]) for existing systematic reviews and relevant methodological research that address these questions. The search-term we used was "guidelines and implementation and systematic review".

We did not conduct systematic reviews ourselves. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Findings

What should WHO do to disseminate and facilitate the uptake of recommendations?

Most research on implementation and dissemination strategies for guidelines have focused on clinical practice guidelines, with change in clinical practice being the primary outcome of interest. An overview of systematic reviews of interventions aimed at changing provider behaviour found that: "In general, passive approaches are generally ineffective and unlikely to result in behaviour change. Most other interventions are effective under some circumstances; none are effective under all circumstances. Promising approaches include educational outreach (for prescribing) and reminders" [9]. A more recent compre-

hensive review of evaluations of the effects of strategies for guideline implementation found that "The majority of interventions observed modest to moderate improvements in care", but there was "considerable variation in the observed effects both within and across interventions"[13].

Few evaluations of interventions to change professional practice have been conducted in low-income countries [14].

Guidance from WHO is often directed towards policy-makers. There is limited research to inform the choice of strategies to improve the uptake of WHO recommendations by policymakers. The findings of systematic reviews of studies of decision-making by health care managers and policymakers have found that factors such as interactions between researchers and health care policy-makers and timing/timeliness appear to increase the prospects for research use among policymakers [15,16].

What should be done at headquarters, by regional offices and in countries?

We did not identify any research findings that could inform the answer to this question.

Discussion

Passive dissemination of guidelines alone is not likely to adequately ensure appropriate uptake of recommendations in most circumstances. However, the conclusion in an extensive review of guidelines implementation strategies was: "There is an imperfect evidence base to support decisions about which guideline dissemination and implementation strategies are likely to be efficient under different circumstances" [13]. Thus, WHO needs to carefully consider the likely benefits and costs of alternative implementation strategies in relationship to specific contexts, and to evaluate the impact of selected strategies.

There are tools available that are designed to assist in the design and evaluation of implementation strategies, such as NorthStar, developed by the EC-funded Research-based continuing education and quality improvement (ReBEQI) project [17]. NorthStar provides a range of information, checklists, examples and tools based on current research on how to best design and evaluate implementation strategies.

Health authorities at national or sub-national levels are better able than WHO to tailor implementation strategies to their specific circumstances. However, they frequently lack capacity and resources to do this. WHO headquarters and regional offices can play an important role in supporting member states in their efforts to implement recom-

mendations by providing tools such as NorthStar, support and coordination of efforts.

Further work

Rigorous evaluations of the effectiveness of strategies for implementing and disseminating recommendations are needed. Given that the use and impact of WHO recommendations is likely to be limited without an active implementation strategy, it is of paramount interest to the organisation to invest in generating the knowledge needed for successful implementation.

Competing interests

AF and ADO work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapy Guidelines.

Authors' contributions

AF prepared the first draft of this review. HJS and ADO contributed to drafting and revising it. All authors read and approved the final manuscript.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

1. Oxman AD, Schünemann H, Fretheim A: **Improving the use of research evidence in guideline development: 16. Evaluation.** *Health Res Policy Syst* 2006, **4**:28.
2. Schünemann H, Fretheim A, Oxman AD: **Improving the use of research evidence in guideline development: 13. Adaptation, applicability and transferability.** *Health Res Policy Syst* 2006, **4**:25.
3. El Arifeen S, Blum LS, Hoque DM, Chowdhury EK, Khan R, Black RE, Victora CG, Bryce J: **Integrated Management of Childhood Illness (IMCI) in Bangladesh: early findings from a cluster-randomised study.** *Lancet* 2004, **364**:1595-1602.
4. Laing R, Hogerzeil H, Ross-Degnan D: **Ten recommendations to improve use of medicines in developing countries.** *Health Policy Plan* 2001, **16**:13-20.
5. Moynihan R, Oxman AD, Lavis J, Paulsen E: **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.
6. Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J: **Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs.** *Int J Qual Health Care* 2003, **15**:31-45.
7. Oxman AD, Fretheim A, Schünemann S: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006, **4**:12.

8. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA: **Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group.** *BMJ* 1998, **317**:465-468.
9. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, Grilli R, Harvey E, Oxman A, O'Brien MA: **Changing provider behavior: an overview of systematic reviews of interventions.** *Med Care* 2001, **39**:112-45.
10. **Getting evidence into practice.** *Effective Health Care* 1999, **5**:1-16.
11. **National Guideline Clearinghouse** Agency for Healthcare Research and Quality 2006 [<http://www.guidelines.gov>].
12. **Guidelines International Network (GIN)** Guidelines International Network 2006 [<http://www.g-i-n.net>].
13. 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**:iii-72.
14. Haines A, Kuruvilla S, Borchert M: **Bridging the implementation gap between knowledge and action for health.** *Bull World Health Organ* 2004, **82**:724-731 [<http://www.who.int/bulletin/volumes/82/10/724.pdf>].
15. Innvaer S, Vist G, Trommald M, Oxman A: **Health policy-makers' perceptions of their use of evidence: a systematic review.** *J Health Serv Res Policy* 2002, **7**:239-244.
16. 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.
17. **Research-based continuing education and quality improvement** www.rebeqi.org 2006 [<http://www.rebeqi.org>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Review

Improving the use of research evidence in guideline development: 16. Evaluation

Andrew D Oxman*¹, Holger J Schünemann² and Atle Fretheim¹

Address: ¹Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and ²INFORMA, S.C. Epidemiologia, Istituto Regina Elena, Via Elio Chianesi 53, 00144 Rome, Italy

Email: Andrew D Oxman* - oxman@online.no; Holger J Schünemann - hjs@buffalo.edu; Atle Fretheim - atle.fretheim@nokc.no

* Corresponding author

Published: 08 December 2006

Received: 07 April 2006

Health Research Policy and Systems 2006, **4**:28 doi:10.1186/1478-4505-4-28

Accepted: 08 December 2006

This article is available from: <http://www.health-policy-systems.com/content/4/1/28>

© 2006 Oxman et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the last of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Objectives: We reviewed the literature on evaluating guidelines and recommendations, including their quality, whether they are likely to be up-to-date, and their implementation. We also considered the role of guideline developers in undertaking evaluations that are needed to inform recommendations.

Methods: We searched PubMed and three databases of methodological studies for existing systematic reviews and relevant methodological research. We did not conduct systematic reviews ourselves. Our conclusions are based on the available evidence, consideration of what WHO and other organisations are doing and logical arguments.

Key questions and answers: Our answers to these questions were informed by a review of instruments for evaluating guidelines, several studies of the need for updating guidelines, discussions of the pros and cons of different research designs for evaluating the implementation of guidelines, and consideration of the use of uncertainties identified in systematic reviews to set research priorities.

How should the quality of guidelines or recommendations be appraised?

- WHO should put into place processes to ensure that both internal and external review of guidelines is undertaken routinely.
- A checklist, such as the AGREE instrument, should be used.
- The checklist should be adapted and tested to ensure that it is suitable to the broad range of recommendations that WHO produces, including public health and health policy recommendations, and that it includes questions about equity and other items that are particularly important for WHO guidelines.

When should guidelines or recommendations be updated?

- Processes should be put into place to ensure that guidelines are monitored routinely to determine if they are in need of updating.
- People who are familiar with the topic, such as Cochrane review groups, should do focused, routine searches for new research that would require revision of the guideline.

- Periodic review of guidelines by experts not involved in developing the guidelines should also be considered.
- Consideration should be given to establishing guideline panels that are ongoing, to facilitate routine updating, with members serving fixed periods with a rotating membership.

How should the impact of guidelines or recommendations be evaluated?

- WHO headquarters and regional offices should support member states and those responsible for policy decisions and implementation to evaluate the impact of their decisions and actions by providing advice regarding impact assessment, practical support and coordination of efforts.
- Before-after evaluations should be used cautiously and when there are important uncertainties regarding the effects of a policy or its implementation, randomised evaluations should be used when possible.

What responsibility should WHO take for ensuring that important uncertainties are addressed by future research when the evidence needed to inform recommendations is lacking?

- Guideline panels should routinely identify important uncertainties and research priorities. This source of potential priorities for research should be used systematically to inform priority-setting processes for global research.
-

Background

The World Health Organization (WHO), like many other organisations around the world, has recognised the need to use more rigorous processes to ensure that health care recommendations are informed by the best available research evidence. This is the last of a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on Health Research to WHO on how to achieve this.

Providing technical advice to its member states is a core function of the World Health Organization (WHO). Ensuring the quality of the advice that is given is an inherent responsibility of WHO. In this paper we address the following questions related to evaluation of guidelines and their implementation:

- How should the quality of guidelines or recommendations be appraised?
- When should guidelines or recommendations be updated?
- How should the impact of guidelines or recommendations be evaluated?
- What responsibility should WHO take for ensuring that important uncertainties are addressed by future research when the evidence needed to inform recommendations is lacking?

Related questions regarding updating of systematic reviews and implementation are addressed in other articles in this series [1,2].

What WHO is doing now

The Guidelines for WHO Guidelines suggest that draft guidelines should be subjected to a self-test by the technical development and the steering/liaison groups using a checklist (Table 1) [3].

However, the checklist is not being used and most guidelines appear to be deficient [4]. An unpublished, in house review of WHO guidelines using the AGREE appraisal instrument [5] found that the vast majority of guidelines did not meet most of the AGREE criteria [6]. Although draft guidelines are often sent for peer review, and the Guidelines for WHO Guidelines suggests external review, this is not always done. WHO's Regulations for Expert Committees, in fact, require that "*The expert committee shall draw up and approve its report before the closure of its meeting.*" [7]

We are not aware of any assessments of the extent to which WHO guidelines, recommendations or policies are kept up to date or any policies for withdrawing ones that are out of date. The Guidelines for WHO Guidelines offers the following encouragement for undertaking rigorous studies to provide missing evidence, but we are not aware of any other policies linking important uncertainties in guidelines to WHO's priorities for research: "*Sometimes it will be necessary to issue guidelines where no rigorous studies exist, based on the best available evidence. But after issuance of such guidelines, the opportunity could be taken to undertake rigorous studies to provide missing evidence and to evaluate the effectiveness or impact of the guidelines in the actual settings where they are intended to be used. This would allow them to be revised or updated if needed.*"

What other organisations are doing

In a survey of 152 organizations that produce guidelines, technology assessments, or support the use of research in

Table 1: Checklist for WHO Treatment Guidelines (from Guidelines for WHO Guidelines [3]).

Yes	Questions	Reference Points
Origin		
1	Are the Cluster and Department issuing the guidelines clearly identified?	Introduction
Objective, target audience		
2	Does the guideline list its objectives, including the patient categories and situation(s) for which the guidelines are intended?	Introduction
3	Does the guideline describe the professional groups to which it is addressed?	Introduction
Guideline Development Group		
4	Does the Guideline Development Group include all relevant professional groups, public health experts and end users, including individuals from geographic areas where the guidelines will be applied?	List of members of the guideline development group
5	Does the Group include methodological experts in fields such as search methodology, critical appraisal and cost-effectiveness analysis?	List of members of the guideline development group
Conflict of interest		
6	Are all funding sources named, and is there no conflict of interest?	List of funding sources
7	Have all members of the Guideline Development Group and external reviewers declared their interests, and have these interests been recorded in the guideline document?	Annex on documentation of process
8	Does the document describe the method used to minimize any undue influence on the Guideline Development Group and the external reviewers?	Annex on documentation of process
Evidence		
9	Was there a systematic comprehensive search for evidence, and has the search strategy been recorded in the guideline?	Annex on documentation of process
10	Has the strength and quality of the evidence on effectiveness been graded?	Annex on documentation of process; evidence table
11a	What percent of recommendations are evidence-based?*	Summary of recommendations
11b	Are the recommendations which are not evidence-based explicitly labeled as "expert opinion" based?	Summary of recommendations
12	Is there explicit consideration of other issues, such as safety and potential misuse in a variety of settings?	Annex on documentation of process; evidence table
13	Is there explicit consideration of issues of cost effectiveness?	Annex on documentation of process; evidence table
14	Is the strength of the recommendation linked to the evidence?	Summary of Recommendations
15	Do the recommendations take into account potential resource constraints?	Implementation issues
Review		
16	Were the comments by the external peer review adequately addressed?	Annex on documentation of process
17	Did all members of the Guideline Development Group approve the final document?	Annex on documentation of process
18	Did all members of the Steering Group approve the final document?	Annex on documentation of process
19	Is there a plan for reviewing new evidence and updating the guideline?	Introduction
Presentation, clarity		
20	Are the recommendations clearly formulated?	Summary of Recommendations
21	Does the guideline identify and advise on ineffective practices?	Summary of Recommendations
Implementation plan		
22	Is there a plan for dissemination and local adaptation of the guideline?	Companion document
23	Are funds available for dissemination and local adaptation for the guideline?	Companion document
24	Are there suggested criteria for monitoring the use in intended settings?	Implementation Issues

This Checklist is intended for the following purposes: (1) As a guide for developing or updating WHO treatment Guidelines. (2) As a check-list for Executive and Regional Directors when giving final approval for publication. To qualify for publication and inclusion in the WHO database of treatment guidelines, a tick mark signifying YES must be placed beside all the 24 criteria, except 11a.

*These are recommendations based on information other than expert opinion.

developing health policy a large majority reported using both an internal review process (80%) and external review by experts (82%) [8]. Only 44% reported external review by target users (58% of guideline producers), 43% reported comparing their products with products or input from other groups, and 31% reported using pilot testing.

Fifty-two percent of the units that produced guidelines reporting updating them regularly and 45% reported updating irregularly. Thirty-five percent reported collecting data about uptake systematically, and 32% reported systematically evaluating the usefulness or impact of their guidelines in some other way.

In another survey of 18 prominent guidelines development programs, more than half reported monitoring or evaluating the effects of at least some guidelines. All reported using some type of quality system for good guideline development. Eleven used both external and internal review, six used external review only, and one internal review only. In addition seven compared their guidelines with guidelines from other groups and three used pilot testing. All reported updating their guidelines at least occasionally, although only half had formal update procedures.

Methods

The methods used to prepare this review are described in the introduction to this series [9]. Briefly, the key questions addressed in this paper were vetted amongst the authors and the ACHR Subcommittee on the Use of Research Evidence (SURE). We did not conduct a full systematic review. We searched PubMed and three databases of methodological studies (the Cochrane Methodology Register, the US National Guideline Clearinghouse, and the Guidelines International Network for existing systematic reviews and relevant methodological research that address these questions. The answers to the questions are our conclusions based on the available evidence, consideration of what WHO and other organisations are doing, and logical arguments.

For this review we used articles that we had previously identified, including a review of clinical practice guideline appraisal instruments [10], to locate related articles in PubMed; we searched the National Guidelines Clearinghouse annotated bibliography using the category guideline evaluation with the terms appraisal or impact, and for all categories using updating; and we checked the reference lists of retrieved articles. We also searched for and scanned reviews of methods for setting research priorities that were linked to guidelines development programs by searching PubMed for reviews of research priorities, for articles that addressed both research priorities and practice guidelines, and by searching the Web using Google for sites that addressed methods for setting priorities for research and global research priorities.

Findings

How should the quality of guidelines or recommendations be appraised?

Graham and colleagues in a systematic review of instruments for assessing the quality of clinical practice guidelines found 13 instruments published up to 1999 [10]. All instruments were developed after 1992 and contained 8 to 142 questions or statements. Only the Cluzeau instrument, which formed the basis of the AGREE instrument [11,12] included at least one item for each of the 10 attributes that the authors identified across instruments.

This instrument and that of Shaneyfelt and colleagues [13]. were the only instruments that had been validated. They concluded that there was insufficient evidence to support the exclusive use of any one instrument, although the AGREE instrument has received the most evaluation. Vlayen and colleagues updated the review by Graham and colleagues up to 2003 [14]. They found 24 different appraisal tools. The 24 instruments included a total of 469 questions that they also grouped into 10 dimensions: validity, reliability/reproducibility, clinical applicability, clinical flexibility, multidisciplinary process, clarity, scheduled review, dissemination, implementation, and evaluation. They found three instruments that addressed all 10 dimensions and three additional instruments based on the Cluzeau instrument, one of which, the AGREE instrument, was the only one to have been validated. They found that the AGREE instrument was a validated, easy-to-use, and transparent instrument, which was internationally developed and widely accepted, but noted two limitations that they considered important: although it can be used to compare clinical practice guidelines, it does not set a threshold to classify them as good or bad, and it does not assess the quality of the evidence supporting the recommendations.

The AGREE instrument was developed through a process of item generation, selection and scaling, field-testing and refinement [5]. The final version of the instrument contained 23 items grouped into six domains: scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, applicability, and editorial independence.

When should guidelines or recommendations be updated?

Shekelle and colleagues, based on a review of 17 guidelines published by AHRQ, estimated that no more than 90% were still valid after 3.6 years and that about half the guidelines were outdated in 5.8 years. They recommend that guidelines should be reassessed every three years, based on the lower 95% confidence interval for their estimate of when one of ten guidelines would no longer be up-to-date. They suggest several ways of expeditiously assessing the need for updating guidelines including conducting limited searches by groups that are familiar with the topic, such as Cochrane review groups, focusing searches on research that the guidelines panel considered would play a pivotal role in requiring revision of the guideline, periodic review of the guidelines by experts not involved in developing the guidelines, and considering guidelines development as an ongoing process, rather than a discreet event, with members of guideline panels serving fixed periods with a rotating membership.

Garthlehner and colleagues compared the approach suggested by Shekelle and colleagues of a limited search using

review articles, commentaries and editorials, to a conventional process using typical systematic review methods in terms of comprehensiveness and effort [15]. They applied both approaches independently to assess the need to update six topics from the 1996 Guide to Clinical Preventive Services from the US Preventive Services Task Force [16]. They found that although the limited search approach identified fewer eligible studies than the traditional approach, none of the studies missed was rated as important by task force members acting as liaisons to the project with respect to whether the topic required an update. On average, this approach produced substantially fewer citations to review than the traditional approach. The effort involved and potential time saving depended largely on the scope of the topic. They found that involving experts in assessing how current the guidelines were was not helpful, in contrast to Shekelle and colleagues.

Johnston and colleagues found that an updating strategy for cancer practice guidelines identified 80 pieces of new evidence over a one-year period relating to 17 of 20 guidelines [17]. They found on average four pieces of new evidence per guideline, but there was considerable variation across the guidelines. Of the 80 pieces, 19 contributed to modifications of clinical recommendations in six practice guidelines, whereas the remaining evidence supported the original recommendations. Their updating process yielded important findings, but was resource intensive. They found that it would be possible to reduce the scope of the sources searched routinely to MEDLINE, the Cochrane Library and meeting proceedings.

The findings of these three studies of the need to update guidelines is consistent with findings from studies of the need to update systematic reviews, which generally support the conclusion that in situations where time or resources are limited, thorough quality assessments should likely take precedence over extensive literature searches [1].

How should the impact of guidelines or recommendations be evaluated?

Strategies ranging from passive dissemination to intensive, complex interventions have been used to implement guidelines and a range of study designs has been used to evaluate the impact of these strategies using a range of outcome measures [18,19]. Passive strategies have often not been effective, however there is limited evidence to support decisions about which guideline dissemination and implementation strategies are likely to be efficient under different circumstances [2,18].

Study designs that can be used to evaluate the impact of guidelines include randomised designs, particularly cluster randomised trials, a range of observational study

designs, including interrupted time series analyses, controlled before-after studies and uncontrolled before-after studies [20]. The advantage of using randomised designs for impact assessments is that they give greater confidence that the measured impact of a program is attributable to whatever implementation strategy was used and not to some other factor [21-23]. It is generally not possible to predict differences in the size, or even the direction, of estimates of treatment effects for the same intervention when it is generated in randomized and non-randomized studies [22]. There have been similar findings for impact evaluations of development programs [21] and implementation strategies. For example, a systematic review of continuous quality improvement found improvements in 41 of 43 single site before-after studies and most of 13 multi site before-after studies, but no improvements in three randomised trials [24].

A wide variety of techniques to gather data have been used singly or in combinations, including questionnaires, interviews, observation, audit and using routinely collected data. Self-report may not be consistent with more objective measures of practice. Collecting reliable data in low and middle-income countries (LMIC) can be a major challenge, where available records and routinely collected data may be lacking. We did not find any systematic reviews of strategies for collecting data for impact evaluations in LMIC.

What responsibility should WHO take for ensuring that important uncertainties are addressed by future research when the evidence needed to inform recommendations is lacking?

Priority-setting exercises for global health research have used various methods and processes [25]. We have not found examples of priority setting programs based on important uncertainties identified in guidelines. A number of exercises have, however, used systematic reviews to inform priority-setting processes [26,27]. A comparison of four sources of potential priorities for the NHS Health Technology Assessment Programme found that a widespread consultation of healthcare commissioners, providers and consumers was the largest source of suggestions, but the success rate of this source, in terms of being commissioned, was low. Research recommendations from systematic reviews provided the second largest source of priorities and the best success rate of all sources.

Discussion

There are at least 24 different instruments available for assessing the quality of clinical practice guidelines. We did not find similar tools developed for assessing the quality of public health or health policy recommendations, although the domains that are addressed by clinical prac-

tice guidelines appraisal instruments are applicable to public health and health policy recommendations.

Up to now, self-assessment of guidelines using the Guidelines for WHO Guidelines checklist has not been successful. Moreover, most guidelines programmes rely on external review, as well as internal review. WHO should put into place processes to ensure that both internal and external review of guidelines is undertaken routinely using appropriate criteria.

Processes should also be put into place to ensure that guidelines are monitored routinely to determine if they are in need of updating. To ensure that this is done as expeditiously as possible, people who are familiar with the topic, such as Cochrane review groups, should conduct limited searches routinely. Guideline panels should identify research that would require revision of the guideline and searches should focus particularly on this research. Periodic review of guidelines by experts not involved in developing the guidelines should also be considered, and consideration should be given to establishing guideline panels that are ongoing with members serving fixed periods with a rotating membership.

Recommendations may need to be adapted to specific settings, can only be implemented in specific settings, and their impact can only be assessed in specific settings. WHO headquarters and regional offices, however, should support member states and those responsible for deciding and implementing policies to evaluate the impact of their policies by providing advice regarding impact assessment, practical support and coordination of efforts. Before-after evaluations should be used cautiously, if at all, and when there are important uncertainties regarding the effects of a policy or its implementation, randomised evaluations should be used when possible.

Guideline panels should routinely identify important uncertainties and research priorities. This source of potential priorities for research should be used systematically to inform priority-setting processes for global research.

Further work

Work is needed to ensure that the AGREE instrument or a similar instrument is suitable for assessing the broad range of guidelines, recommendations and policies that WHO produces. In particular, its suitability for assessing public health and health policy recommendations should be assessed. Additional items should also be added to address concerns about equity, which are not currently addressed in the AGREE instrument, and considerations that are specific to guidelines that are developed internationally rather than in a specific country or setting. Work is also needed on developing practical methods to collect

reliable data that can be used in impact evaluations in LMIC.

Competing interests

ADO and AF work for the Norwegian Knowledge Centre for the Health Services, an agency funded by the Norwegian government that produces systematic reviews and health technology assessments. All three authors are contributors to the Cochrane Collaboration. ADO and HJS are members of the GRADE Working Group. HJS is documents editor and chair of the documents development and implementation committee for the American Thoracic Society and senior editor of the American College of Chest Physicians' Antithrombotic and Thrombolytic Therapeutic Guidelines.

Authors' contributions

ADO prepared the first draft of this review. HJS and AF contributed to drafting and revising it.

Acknowledgements

We gratefully acknowledge WHO and the Norwegian Knowledge Centre for the Health Services for financial support, the other authors of these reviews for their contributions, the other members of the ACHR SURE, and all of those who have provided us with feedback on earlier versions of this paper and the others in this series.

References

- Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006.
- Fretheim A, Schünemann HJ, Oxman AD: **Improving the Use of Research Evidence in Guideline Development: 15. Disseminating and implementing guidelines.** *Health Res Policy Syst* 2006.
- Global Programme on Evidence for Health Policy: **Guidelines for WHO Guidelines.** Geneva: World Health Organisation, 2003 (EIP/GPE/EQC/2003.1).
- Panisett U: **A review of WHO recommendations published in 2005.**
- 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.
- Oxman AD, Lavis J, Fretheim A: **The use of research evidence in WHO recommendations.**
- Regulations for Expert Advisory Panels and Committees: **In World Health Organisation. Basic Documents, 2003; 101-9. Text adopted by the Thirty-fifth World Health Assembly (resolution WHA35.10), in replacement of the regulations adopted by the Fourth World Health Assembly, 1951. Amendments were adopted at the Forty-fifth, Forty-ninth, Fifty-third and Fifty-fifth World Health Assemblies (decision WHA45(10), resolution WHA49.29, resolution WHA53.8 and resolution WHA55.24, respectively.**
- Moynihhan R, Oxman AD, Lavis JN, Paulsen E: **Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier.** In *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.
- Oxman AD, Fretheim A, Schünemann HJ: **Improving the use of research evidence in guideline development: introduction.** *Health Res Policy Syst* 2006.
- Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM: **A comparison of clinical practice guideline appraisal instruments.** *Int J Technol Assess Health Care* 2000, **16**:1024-38.

11. The 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.
12. AGREE Instrument: **Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration.** [<http://www.agreecollaboration.org/instrument/>].
13. 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.
14. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D: **A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit.** *Int J Qual Health Care* 2005, **17**:235-42.
15. Gartlehner G, West SL, Lohr KN, Kahwati L, Johnson JG, Harris RP, et al.: **Assessing the need to update prevention guidelines: a comparison of two methods.** *Int J Qual Health Care* 2004, **16**:399-406.
16. US Preventive Services Task Force: **Guide to Clinical Preventive Services.** 2nd edition. Alexandria, VA: International Medical Publishing; 1996.
17. Johnston ME, Brouwers MC, Browman GP: **Keeping cancer guidelines current: results of a comprehensive prospective literature monitoring strategy for twenty clinical practice guidelines.** *Int J Technol Assess Health Care* 2003, **19**:644-55.
18. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, et al.: **Effectiveness and efficiency of guideline dissemination and implementation strategies.** *Health Technol Assess* 2004, **8**.
19. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, et al.: **Changing provider behavior: an overview of systematic reviews of interventions.** *Med Care* 2001, **39**:112-45.
20. Eccles M, Grimshaw J, Campbell M, Ramsay C: **Research designs for studies evaluating the effectiveness of change and improvement strategies.** *Qual Saf Health Care* 2003, **12**:47-52.
21. Savedoff W, Levine R, Birdsall N: **hen We Will Ever Learn? Closing the Evaluation Gap. Consultation Draft.** *Center for Global Development* [<http://www.cgdev.org/section/initiatives/active/evalgap/>]. (Accessed 16 March 2006).
22. Oxman AD, Schünemann HJ, Fretheim A: **Improving the Use of Research Evidence in Guideline Development: 7. Deciding what evidence to include.** *Health Res Policy Syst* 2006.
23. Kleijnen J, Gøtzsche P, Kunz RA, Oxman AD, Chalmers I: **So what's so special about randomisation.** In *Non-Random Reflections on Health Care Research: On the 25th Anniversary Of Archie Cochrane's Effectiveness and Efficiency* Edited by: Chalmers I, Maynard A. London: BMJ Publishers; 1997:93-106.
24. Shortell SM, Bennett CL, Byck GR: **Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress.** *Milbank Q* 1998, **76**:593-624.
25. Global Forum for Health Research: **Priority setting in health research.** *The 10/90 Report on Health Research 2003–2004* [http://www.globalforumhealth.org/filesupld/1090_report_03_04/109004_chap_4.pdf]. (Accessed 16 March 2006).
26. The James Lind Alliance: **Confronting important uncertainties about the effects of treatments.** *Health and Social Campaigners News* 2004:3-10 [http://www.jameslindlibrary.org/pdf/effects_of_treatments.pdf]. (Accessed 16 March 2006).
27. Chase D, Milne R, Stein K, Stevens A: **What are the relative merits of the sources used to identify potential research priorities for the NHS HTA programme?** *Int J Technol Assess Health Care* 2000, **16**:743-50.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

