

Guidance for Evidence-Informed Policies about Health Systems: Assessing How Much Confidence to Place in the Research Evidence

Simon Lewin^{1*}, Xavier Bosch-Capblanch², Sandy Oliver³, Elie A. Akl⁴, Gunn E. Vist⁵, John N. Lavis⁶, Davina Ghersi⁷, John-Arne Røttingen⁸, Peter Steinmann⁹, Metin Gulmezoglu¹⁰, Peter Tugwell¹¹, Fadi El-Jardali¹², Andy Haines¹³

1 Norwegian Knowledge Centre for the Health Services, Oslo, Norway, and Health Systems Research Unit, Medical Research Council of South Africa, Cape Town, South Africa, **2** Swiss Tropical and Public Health Institute, Basel, and University of Basel, Basel, Switzerland, **3** EPPI-Centre, Social Science Research Unit, Institute of Education, London, United Kingdom, **4** Department of Medicine, State University of New York at Buffalo, Buffalo, New York, United States of America, and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, **5** Norwegian Knowledge Centre for the Health Services, Oslo, Norway, **6** McMaster Health Forum, Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, Hamilton, Ontario, Canada, **7** Department of Research Policy and Cooperation, Information, Evidence and Research, World Health Organization, Geneva, Switzerland, **8** Norwegian Knowledge Centre for the Health Services, Oslo, Norway, **9** Swiss Tropical and Public Health Institute, Basel, Switzerland, **10** UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization, Switzerland, **11** Department of Medicine, Institute of Population Health, and Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada, **12** Department of Health Policy and Management, American University of Beirut, Beirut, Lebanon, and McMaster Health Forum, McMaster University, Hamilton, Ontario, Canada, **13** Departments of Social and Environmental Health Research and of Nutrition and Public Health Research, London School of Hygiene and Tropical Medicine, London, United Kingdom

This is one paper in a three-part series that sets out how evidence should be translated into guidance to inform policies on health systems and improve the delivery of clinical and public health interventions.

Introduction

Health systems interventions establish or modify governance (e.g., licensing of professionals), financial (e.g., health insurance mechanisms) and delivery (e.g., by whom care is provided) arrangements, and implementation strategies (e.g., strategies to change health provider behaviours) within health systems (which consist of “all organisations, people and actions whose primary intent is to promote, restore or maintain health”; see Box S1 for definitions of the terms used in this article). The focus of these interventions is to strengthen health systems in their own right or to get cost-effective programmes and technologies (e.g., drugs and vaccines) to those who need them. Decisions regarding health systems strengthening, including the development of recommendations by policy makers, require evidence on the effectiveness of these interventions, as well as many other forms of evidence. For example, in

assessing potential policy options, reviews of economic evaluations and of qualitative studies of stakeholders’ views regarding these options might be important (Table S1). Such evidence helps to address questions such as the cost-effectiveness of these options and which options are seen as appropriate by stakeholders.

Assessing how much confidence to place in the types of evidence available on health systems interventions is a key component in informing judgements regarding the use of such interventions for health systems strengthening (Box 1). This paper, which is the third of a three-part series on health systems guidance [1,2], aims to:

- Illustrate a range of tools available to assess the different types of evidence needed to support different steps in the policy-making process;
- Discuss the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach to assessing confidence in estimates of effects (“quality of evidence”) and to grading the strength of recommendations on policy options for health systems interventions;
- Discuss factors that are important when developing recommendations on policy options regarding health systems interventions.

Citation: Lewin S, Bosch-Capblanch X, Oliver S, Akl EA, Vist GE, et al. (2012) Guidance for Evidence-Informed Policies about Health Systems: Assessing How Much Confidence to Place in the Research Evidence. *PLoS Med* 9(3): e1001187. doi:10.1371/journal.pmed.1001187

Published: March 20, 2012

Copyright: © 2012 Lewin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The development of the Handbook was supported by a grant to the WHO by the Rockefeller Foundation, which had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The paper represents the views of the authors and neither WHO nor the Rockefeller Foundation.

Competing Interests: The Swiss Tropical and Public Health Institute and the Norwegian Knowledge Centre for the Health Services received funds from the WHO for the contributions of DD, PS, LB, SL, and XBC to developing a Handbook to produce health systems guidance, and some of this work is reported in this article. DG is a member of the *PLoS Medicine* Editorial Board. EAA and GEV are members of the GRADE Working Group. All other authors declare no competing interests.

Abbreviations: GRADE, Grading of Recommendations Assessment, Development and Evaluation; WHO, World Health Organization

* E-mail: simon.lewin@nokc.no

Provenance: Not commissioned; externally peer reviewed.

The Policy Forum allows health policy makers around the world to discuss challenges and opportunities for improving health care in their societies.

Summary Points

- Assessing how much confidence to place in different types of research evidence is key to informing judgements regarding policy options to address health systems problems.
- Systematic and transparent approaches to such assessments are particularly important given the complexity of many health systems interventions.
- Useful tools are available to assess how much confidence to place in the different types of research evidence needed to support different steps in the policy-making process; those for assessing evidence of effectiveness are most developed.
- Tools need to be developed to assist judgements regarding evidence from systematic reviews on other key factors such as the acceptability of policy options to stakeholders, implementation feasibility, and equity.
- Research is also needed on ways to develop, structure, and present policy options within global health systems guidance.
- This is the third paper in a three-part series in *PLoS Medicine* on health systems guidance.

The first paper in this series makes a case for developing guidance to inform decisions on health systems questions and explores challenges in producing such guidance and how these might be addressed [1]. The second paper explores the links between guidance development and policy development at global and national levels, and examines the range of factors that can influence policy development [2].

In this paper, which like the other two papers is based on discussions of the Task Force on Developing Health Systems Guidance (Box 2; [1,2]), we focus particularly on the GRADE approach, which provides a transparent and systematic approach to rating the quality of evidence and grading the strength of recommendations [3].

Box 1. Reasons Why It Is Important to Assess How Much Confidence Can Be Placed in Evidence to Guide Decisions on Health Systems Strengthening

- Users of such evidence almost always draw implicit or explicit conclusions regarding how much confidence to place in evidence
- Such evidence may inform judgements regarding recommendations, including the strength of such recommendations
- Systematic and explicit approaches can be useful in:
 - facilitating critical appraisal of evidence
 - protecting against bias
 - clarifying implementation issues
 - resolving disagreements among stakeholders
 - communicating information regarding the evidence, the judgements made, and recommendations drawn from it
- Systematic and explicit approaches are particularly important given the complexity of many health systems interventions. These interventions may be complex in terms of the number of discrete, active components and the interactions between them; the number of behaviours to which the intervention is directed; the number of organisational levels targeted by the intervention; the degree of flexibility or tailoring permitted in intervention implementation; the level of skill required by those delivering the intervention; and the extent of context dependency [48,49]. Given the complexity of these interventions, deciding on the contextual relevance of evidence is a crucial component.

Source: adapted from [30].

Tools to Assess the Evidence Needed to Support the Policy-Making Process for Health Systems Strengthening

Well-conducted systematic reviews [4] can be used to identify the best available evidence to inform judgements about the effects of policy options and to inform other key steps within the policy-making process (Table S1). As discussed in the second paper of this series [2], users need to be able to assess the quality of evidence presented in such reviews in relation to each step of the policy-making process. For example, when defining the problem and the need for intervention, tools are required to assess the confidence we can place in evidence from reviews of studies highlighting different ways of conceptualising the problem (e.g., reviews of studies of people's experiences of the problem) [5]. When assessing potential policy options, tools are needed to assess the confidence that can be placed in, for example, studies assessing impact (e.g., reviews of effectiveness studies). Similarly, when identifying implementation considerations, tools are required to assess the confidence that can be placed in reviews of factors affecting implementation.

Many tools are available to assess the risk of bias in individual studies of the effects of interventions [6] and to appraise individual qualitative studies [7]. Tools are also available to assess the quality of evidence synthesised in systematic reviews [6]. Such tools need to be appropriate to the types of studies included in the review and generic enough to be applicable across a range of questions, and must allow meaningful conclusions to be drawn regarding the quality of the included evidence. Judgements on how much confidence can be placed in the evidence from a review need to be distinguished from judgements about how well the review was conducted (i.e., its reliability). Tools have been developed to assess the reporting of systematic reviews and meta-analyses (e.g., the PRISMA checklist [8]) and to assess their methodological quality or reliability (e.g., the SUPPORT tools [9] and AMSTAR [10]) (Table 1). However, we focus here on tools to assess how much confidence can be placed in the evidence identified and presented in those reviews.

Assessing How Much Confidence to Place in the Findings of Reviews of the Effects of Policy Options

Tools to assess how much confidence to place in review findings are most devel-

Box 2. The Task Force on Developing Health Systems Guidance

- The Task Force on Developing Health Systems Guidance was established in 2009 by the World Health Organization (WHO) to improve its response to requests for guidance on health systems.
- The Task Force consisted of 20 members selected by WHO for their expertise in the field of health systems research and implementation.
- Through a series of face-to-face and virtual meetings, the Task Force provided input to and oversight of the development of a *Handbook for Developing Health Systems Guidance* and to the identification of broader issues that warranted further dialogue and debate [39].
- As part of this process, the Task Force and the Handbook developers reviewed approaches to developing clinical guidelines and the instruments used for clinical guideline development.
- The Task Force suggested ways in which some of these approaches and instruments could be adapted for use in the development of health systems guidance and indicated where there were important differences between these approaches.
- The writing group for this paper further considered the issues raised in these discussions and produced a first draft of the manuscript for comment by the Task Force.
- This paper, and the other two in the series [1,2], were finalised after several iterations of comments by the Task Force and external reviewers.

oped for systematic reviews of evidence on effectiveness. The GRADE approach is one such tool, but many others are available [11]. Within GRADE, the quality of evidence derived from a systematic review is related to the quality of the included studies and to a range of other factors (Table S2). This approach has many strengths (Table S3) and is now used increasingly by international organizations, including the World Health Organization (WHO), the Cochrane Collaboration, and several agencies developing guidelines [3]. (Also see <http://www.gradeworkinggroup.org>.) We will discuss

the application of the GRADE approach to assess the quality of evidence and the strength of recommendations for health systems interventions later.

Assessing How Much Confidence to Place in the Findings of Reviews of Questions Other Than Effects

Tools to assist judgements on how much confidence to place in the findings of reviews of questions other than effects that are relevant to the policy-making process are at an early stage of development. Such questions include stakeholders' values and preferences and the feasibility of interven-

tions (Table S1). For some of these issues, judgements might be informed by systematic reviews of qualitative studies, together with local evidence [12]. Where reviews seek a qualitative answer to understand the nature of a problem, quality appraisal aims to assess the coherence of the resulting explanation, possibly across different contexts. Although quality criteria for individual qualitative research studies commonly consider the methods of each study and the credibility and richness of their findings [7,13,14], thus far tools for assessing the quality of systematic reviews of qualitative research have not considered the credibility and richness of findings. A potential tool for doing this is proposed in Table S4.

Reviews exploring factors affecting the implementation of options might employ mixed methods syntheses (i.e., syntheses of both qualitative and quantitative evidence) such as realist synthesis [15], which explores the explanatory theories implicit in existing programmes or policies, or framework synthesis, which provides a highly structured, deductive approach to data analysis drawing on an existing model or framework (for example, [16–18]). Quality appraisal then focuses on the confidence that can be placed in each conclusion drawn from individual studies [19]. Grading the evidence as a whole can take into account the number and context of the studies contributing to each conclusion, and the appropriateness of their methods for drawing that conclusion (for illustration, see [16]). A single study might refute or qualify a theory, but multiple studies together contribute to strengthening a theory. As yet, there are no tools for appraising how well mixed methods re-

Table 1. Commonly used tools to assess systematic reviews and their findings and to assess clinical guidelines.

<i>Systematic Reviews</i>	
SUPPORT (SUPporting Policy Relevant Reviews and Trials) tool [9]	A tool to assess how much confidence to place in the methodological quality of a systematic review, and designed for reviews of health systems interventions
AMSTAR (A measurement tool for the "assessment of multiple systematic reviews") [10]	A tool designed to assess the methodological quality of a systematic review. This tool has not been designed specifically to assess how much confidence to place in reviews of health systems interventions
GRADE (Grading of Recommendations Assessment, Development and Evaluation) [41]	An approach to assess the quality of evidence
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analyses) [8]	A tool to assess the reporting of systematic reviews and meta-analyses
SUPPORT applicability tool [33] and Wang applicability and transferability tool [34]	Tools to assist in assessing the applicability of the findings of a systematic review to a specific setting
<i>Guidelines</i>	
AGREE (Appraisal of Guidelines for Research and Evaluation) [42]	A tool to evaluate the process of practice guideline development and the quality of reporting
GRADE [41]	An approach to grade the strength of recommendations

doi:10.1371/journal.pmed.1001187.t001

views have synthesised studies to draw conclusions about the advantages or disadvantages of policy options.

Resource use is another key issue, and tools are available to assess the reliability of reviews of economic studies [20]. In addition, GRADE provides guidance on how to incorporate considerations of resource use into recommendations [21].

The GRADE Approach to Assessing Confidence in the Estimates of Effects for Health Systems Interventions

The GRADE approach clearly separates two issues: the quality of the evidence and the strength of recommendations. Quality of evidence is only one of several factors considered when assessing the strength of recommendations.

Within the context of a systematic review, GRADE defines the quality of evidence as the extent to which one can be confident in the estimate of effect. Within the context of guidelines or guidance, GRADE defines the quality of evidence as the confidence that the effect estimate supports a particular recommendation. The degree of confidence is a continuum but, for practical purposes, it is categorised into high, moderate, low, and very low quality (Table S5).

Evidence on the effectiveness of health systems interventions raises a number of challenges that may, in turn, influence assessments of the quality of this evidence and the development of recommendations using the GRADE approach. Firstly, while experimental studies (including pragmatic randomised trials [22]) are feasible for some health systems interventions, for others (particularly those related to governance and financial arrangements), evidence may come mainly from observational studies, including evaluations of national or state-wide programmes [23,24]. Secondly, evaluations of health systems interventions often use clustered designs and these are frequently poorly conducted, analysed, and reported [25,26]. Thirdly, health systems interventions tend to measure proxy outcomes, such as the use of services or the uptake of an incentive. Evidence users need to decide whether there is sufficiently strong evidence of a relationship between the proxy outcome and the desired health outcome. The development of an outcomes framework to assist in assessing interventions (for an example, see [27,28]) may help those developing guidance

decide whether proxy outcomes are sufficient. Finally, poorly described health systems and political systems factors and implementation considerations may make it difficult to develop contextualised recommendations on policy options [29]. GRADE attempts to make the judgements regarding these issues systematic and transparent.

Developing Recommendations on Policy Options for Consideration Regarding Health Systems Interventions

Moving from evidence to recommendations on options for consideration often necessitates the interpretation of factors other than evidence. In most cases, these interpretations require judgments, making it important to be transparent, particularly given that recommendations will sometimes need to consider multiple complex health systems interventions, each with its own assessment of quality of evidence. Another challenge is the additional complexity of assessing the wide range of health system and political system factors that will influence the choice and implementation of options for addressing a health system problem in different settings (see the other papers in this series [1,2]). For example, a health systems problem may involve a wide range of stakeholders, each with views regarding the available options. In addition, health systems interventions may have system-wide effects that vary across settings. Consequently, rather than making a single recommendation as in clinical guidelines, it may be more useful for health systems guidance to set out the evidence and outline a range of options, appropriate to different settings, to address a given health systems problem. These options may, in turn, feed into deliberative or decision-making processes at national or sub-national levels, as discussed later and elsewhere in this series [2].

Tools such as GRADE assist in grading the strength of a recommendation regarding options [30,31] and can be applied to health systems interventions, but may benefit from explicitly including some additional factors (Box 3). Further research is needed to explore the usefulness of these additional factors but, in general, any tool used to guide the development of recommendations should aim to improve transparency by explicitly describing the factors, and their interpretation, that contributed to the development of recommendations.

Within the GRADE approach, recommendations reflect the degree of confidence that the desirable effects of applying a recommendation outweigh the undesirable effects. Specifically, a strong recommendation implies confidence that the desirable effects of applying a recommendation outweigh the undesirable effects, whereas a conditional/qualified/weak recommendation suggests that the desirable effects of applying a recommendation probably outweigh the undesirable effects, but there is uncertainty.

GRADE attempts to make all judgements regarding the factors that are considered in developing recommendations transparent (by documenting these judgments) and systematic (by using the same approach across all the questions being considered by the guideline). Tables 2, S6, and S7 provide illustrations of the application of the GRADE approach to health systems interventions involving delivery and financial arrangements and implementation strategies, respectively, and show how guidance on health systems interventions might outline a range of options appropriate to different settings—an approach on which further research is needed. In common with other grading systems, GRADE does not yet provide guidance on how to assess the level of confidence that can be placed in evidence on “acceptability” or “feasibility”. Conventionally, these judgements have been made by consensus among the guideline panel, which needs to include individuals with expertise and experience relevant to the guideline questions. Further work is needed to develop a formal way of assessing the quality of such evidence.

Challenges in Moving from an Assessment of the Quality of Evidence to Making Recommendations on Policy Options

The move from an assessment of the quality of evidence to making recommendations on policy options involves a number of challenges. Firstly, assessments of the strength of a recommendation may require a detailed understanding of the evidence creation and evaluation process that is not always available. Secondly, categorising recommendations as “strong” and “weak” can raise difficulties. Panels developing global guidance may, for example, be reluctant to make “weak” recommendations in case policy makers fail to respond to such recommendations because they assume they are equivalent to

Box 3. Factors That May Inform Decisions about the Strength of Recommendations Regarding Policy Options

GRADE factors (adapted from [30]):

- Whether there is uncertainty about the balance of benefits versus harms and burdens
- The quality of the evidence from the systematic review (very low, low, moderate, high)
- Whether there is uncertainty or variability in values and preferences among stakeholders
- Whether there is uncertainty about whether the net benefits are worth the costs or about resource use
- Whether there is uncertainty about the feasibility of the intervention (or about local factors that influence the translation of evidence into practice, including equity issues)

Additional factors that it may be useful to consider for health systems interventions:

- Ease of implementation at the systems level, including governance arrangements (e.g., changes needed in regulations), financial arrangements (e.g., the extent to which the options fit with financing models within settings), and implementation strategies (e.g., how to provide the skills and experience needed among implementers or facilitators)
- Socio-political considerations, e.g., how the proposed options relate to existing policies, values within the political system in relation to issues such as equity or privatisation, and economic changes

“no recommendation”. Thirdly, the quality of evidence is typically assessed for two alternative policy options in tools such as GRADE, while many health system (and, indeed, many clinical decisions) involve multiple interventions, which adds to the complexity of interpretation and decision-making and makes it even more important to be transparent.

How might some of these challenges be addressed? Methodological expertise is needed to conduct and interpret systematic reviews and perform assessments using tools such as GRADE. Health systems guidance panels therefore need to be supported by methodology experts. Moreover, the outputs of these tools need to be “translated” into appropriate language and formats to ensure that they can be interpreted and used correctly by the panel. Research is under way within initiatives such as the DECIDE (Developing and Evaluating Communication strategies to support Informed Decisions and practice based on Evidence; <http://www.decide-collaboration.eu>) collaboration on ways of presenting information on GRADE assessments and policy options to policy makers.

Additional Challenges

There are other wider challenges involved in making recommendations on policy options regarding health systems

interventions. Firstly, the tools used to assess the quality of evidence and develop recommendations need to be able to accommodate the wide range of study designs that is used to assess the effectiveness of health system interventions. This is possible within GRADE. Secondly, tools need to be developed to inform judgements on how much confidence to place in the other forms of evidence (e.g., evidence on acceptability) that are needed to develop recommendations regarding health systems interventions [32].

Finally, international standard-setting organisations, such as WHO, have to formulate recommendations that are applicable at a global level. However, as noted earlier, creating global recommendations on health systems questions can be difficult because of important variations in context-specific factors that influence the applicability of interventions at national and sub-national levels [2,33]. An approach that should help to link guidance development at the global level with policy development at the national level is outlined in the second paper of this series [2].

Where it is useful to make recommendations at the global level, those developing guidance may choose to outline policy options rather than a single recommendation. Such options may encompass one or more questions and may be based on the range of interventions considered in rela-

tion to these questions. Health and political systems factors could be taken into account by linking specific options to these factors. For example, the options may describe variations in the intervention content and method of implementation, based on the evaluations that have been conducted in different settings. Further work is needed to explore how policy makers might interpret and select policy options outlined in global guidance. However, one useful approach might be to provide national decision makers with tools to assist them in making recommendations appropriate to their setting. Several such tools are available [33,34] or in development (see <http://www.decide-collaboration.eu/work-packages-strategies>). Importantly, global guidance should always indicate the factors that should be considered to assess the implications of variations in intervention, context, and other conditions. Decision models may be useful in exploring the effects of these variations (for example, [35,36]). Given the often low quality evidence available regarding policy options for health systems problems, it is also likely that in many cases the recommended option(s) will need to be evaluated.

Presenting Evidence Regarding Contextual and Implementation Issues to Guidance Panels and Policy Makers

The best way to communicate evidence on contextual and implementation issues related to health systems and political systems to guidance panels and policy makers to inform their judgements about the strength of recommended options is currently unclear. Related work on summary of findings tables for systematic reviews of effects and evidence summaries for policy makers has illustrated the importance of paying attention to both format and content in developing useful and understandable presentation approaches [37,38]. The *Handbook for Developing Health Systems Guidance* sets out an approach for presenting this type of evidence to stakeholders in a user-friendly evidence profile [39]. Similarly, the second paper in this series describes the wider features of health and political systems that may need to be assessed to inform decision-making [2].

If we want to ensure that guidance panels and policy makers use evidence to inform judgements about the strength of recommended options, more research is needed to develop and test approaches (including visual formats) for presenting the available evidence to such groups. In addition, efforts are needed to build the

Table 2. Example of factors affecting decisions about strength of recommendations—Lay or community health workers to reduce childhood mortality.

Population: Children in high mortality settings Intervention: Lay health workers (LHWs) delivering health promotion, treatment, and referral interventions Comparison: No LHW intervention / usual care Outcome: Childhood mortality		
Key factors—is there uncertainty regarding:	Decision regarding whether there is uncertainty (yes / no)	Explanation of the decision made
Quality of evidence	Yes	The use of LHWs in maternal and child health programmes may lead to fewer deaths among children under five (low quality evidence—GRADE). In addition, the use of LHWs probably leads to an increase in the number of women who breastfeed and to the number of children who have their immunisation schedule up to date (moderate quality evidence for both outcomes—GRADE). These additional outcomes are also related to mortality reduction
Balance of benefits versus harms and burdens	Yes	Potentially important benefits (mortality reduction) but confidence interval also includes harm. Additional evidence on LHWs suggests effectiveness, e.g., LHWs associated with increased uptake of interventions of proven cost-effectiveness (immunisation, breastfeeding)
Acceptability	Yes	<ul style="list-style-type: none"> Some evidence that LHWs acceptable to service users and used widely Varied acceptability to other services providers in different settings (e.g., [43,44])
Resource use	Yes	Potentially large investment needed over long period but alternatives likely to be more costly
Feasibility (or local factors that influence the translation of evidence into practice)	Yes	There may be constraints to scaling up trained LHWs and supporting them, but it is even less feasible to scale up professional cadres. There are a number of well-documented examples of LHW programmes that have been taken to scale for which monitoring has suggested some positive outcomes, e.g., in Ethiopia and Pakistan [45,46]
Recommended options for consideration <i>This assessment of evidence within a wider health system context might result in the following recommended options for consideration:</i> <ul style="list-style-type: none"> Option 1: Where child mortality is high; an infrastructure for LHWs can be developed rapidly; it is unlikely that the numbers of other cadres could be expanded; and this cadre is acceptable to other providers and to service users and has strong political support: <ul style="list-style-type: none"> Strong recommendation to implement LHWs to reduce childhood mortality (i.e., there is confidence that the desirable effects of LHWs delivering interventions to reduce childhood mortality outweigh the undesirable effects). Option 2: Where child mortality is high; LHWs are acceptable to other providers and to service users; but governance and financing mechanisms for LHWs will need to be established, and there is little experience of running such programmes and uncertainty among policy makers: <ul style="list-style-type: none"> Conditional recommendation to implement LHWs to reduce childhood mortality (i.e., the desirable effects of LHWs delivering interventions to reduce childhood mortality probably outweigh the undesirable effects, but there is uncertainty). Option 3: Where child mortality is moderate; an infrastructure for LHWs can be developed rapidly; but there is evidence that the scaling up of LHWs may be challenged by health care professionals: <ul style="list-style-type: none"> Conditional recommendation to implement LHWs to reduce childhood mortality, dependent on the ability to overcome professional opposition (i.e., the desirable effects of LHWs delivering interventions to reduce childhood mortality probably outweigh the undesirable effects, but there is uncertainty). 		

Source: This table draws on evidence from [47].
doi:10.1371/journal.pmed.1001187.t002

capacity of policy makers to use evidence to inform their decisions [2,12,40].

Conclusions

Useful tools are available for grading quality of evidence and strength of recommendations on policy options regarding health systems interventions, but several challenges need to be addressed. Firstly, these tools involve judgements, and these need to be made systematically and transparently. Secondly, for many health systems questions, evidence is still likely to be of low quality. Better quality research in

these areas is needed and would allow guidance panels to have more confidence in the evidence and to issue stronger recommendations. Thirdly, research is needed on ways to develop, structure, and present policy options for consideration within global health systems guidance. These options need to include evidence on health and political system factors and implementation considerations, and tools to assess such evidence need to be refined. Finally, greater attention needs to be given to how guidance on health systems interventions may be implemented at the local level.

Supporting Information

Alternative Language Summary Points S1 Translation of the Summary Points into Spanish by Xavier Bosch-Capblanch
(DOC)

Alternative Language Summary Points S2 Translation of the Summary Points into French by Bruno Clary, William Lenoir, and Lise Beck
(DOC)

Alternative Language Summary Points S3 Translation of the Sum-

mary Points into Portuguese by Bruno Viana
(DOC)

Alternative Language Summary Points S4 Translation of the Summary Points into Arabic by Fadi El-Jardali
(DOC)

Box S1 Definitions of terms used in this paper
(PDF)

Table S1 Types of systematic reviews needed for different steps in the policy-making process and the availability of tools to assess how much confidence can be placed in the evidence presented in these reviews
(PDF)

Table S2 GRADE criteria for assessing the quality of evidence for each important outcome assessed in a systematic review of effects
(PDF)

Table S3 Strengths of the GRADE approach
(PDF)

Table S4 Assessing how much confidence can be placed in the findings

of systematic reviews of qualitative studies
(PDF)

Table S5 Definitions of the quality of evidence categories within GRADE
(PDF)

Table S6 Example of factors affecting decisions about strength of recommendations—Changes in user fees in low- and middle-income countries
(PDF)

Table S7 Example of factors affecting decisions about strength of recommendations—Continuing education programmes for rural health workers to support their retention
(PDF)

Acknowledgments

We acknowledge other members of the Task Force on Developing Health Systems Guidance, who include (with their affiliations at the time when the Task Force was initiated): Edgardo Abalos, Centro Rosarino de Estudios Perinatales (Argentina); Abdul Ghaffar, Alliance for Health Policy and Systems Research, World Health Organization (Switzerland); Timothy

Evans, Assistant Director-General, Information, Evidence and Research, World Health Organization (Switzerland); Regina Kulier, Department of Research Policy and Cooperation, Information, Evidence and Research, World Health Organization (Switzerland); Pierre Ongolo-Zogo, Centre for the Development of Best Practices in Health, Yaoundé Central Hospital, and University of Yaoundé, Yaoundé (Cameroon); Tikki Pang, Department of Research Policy and Cooperation, Information, Evidence and Research, World Health Organization (Switzerland); Ulysses Panisset, Department of Research Policy and Cooperation, Information, Evidence and Research, World Health Organization (Switzerland).

Thanks also to the following for their thoughtful input: Claire Glenton and Andy Oxman, Norwegian Knowledge Centre for the Health Services (Norway).

Author Contributions

Conceived and designed the experiments: SL XBC SO EAA GEV JL DG JAR PS MG PT FEJ AH. Wrote the first draft of the manuscript: SL. Contributed to the writing of the manuscript: XBC SO EAA GEV JL DG JAR PS MG PT FEJ AH. ICMJE criteria for authorship read and met: SL XBC SO EAA GEV JL DG JAR PS MG PT FEJ AH. Agree with manuscript results and conclusions: SL XBC SO EAA GEV JL DG JAR PS MG PT FEJ AH.

References

1. Bosch-Capblanch X, Lavis JN, Lewin S, Atun R, Rottingen JA, et al. (2012) Guidance for evidence-informed decisions about health systems: rationale for and challenges of guidance development. *PLoS Med* 9: e1001185. doi:10.1371/journal.pmed.1001185.
2. Lavis JN, Rottingen JA, Bosch-Capblanch X, Atun R, El-Jardali F, et al. (2012) Guidance for evidence-informed policies about health systems: linking guidance development to policy development. *PLoS Med* 9: e1001187. doi:10.1371/journal.pmed.1001187.
3. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, et al. (2011) GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64: 383–394.
4. Lavis JN, Oxman AD, Lewin S, Fretheim A (2009) SUPPORT tools for evidence-informed health policymaking (STP). *Health Res Policy Syst* 7 Suppl 1: 11.
5. Lavis JN, Wilson MG, Oxman AD, Grimshaw J, Lewin S, et al. (2009) SUPPORT tools for evidence-informed health policymaking (STP) 5: using research evidence to frame options to address a problem. *Health Res Policy Syst* 7 Suppl 1: S5.
6. West S, King V, Carey TS, Lohr KN, McKoy N, et al. (2002) Systems to rate the strength of scientific evidence. Evidence report/technology assessment No. 47 (prepared by the Research Triangle Institute—University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ publication no. 02-E016.
7. Cohen DJ, Crabtree BF (2008) Evaluative criteria for qualitative research in health care: controversies and recommendations. *Ann Fam Med* 6: 331–339.
8. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 62: e1–e34.
9. Lewin S, Oxman AD, Lavis JN, Fretheim A (2009) SUPPORT tools for evidence-informed health policymaking (STP) 8: deciding how much confidence to place in a systematic review. *Health Res Policy Syst* 7 Suppl 1: S8.
10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, et al. (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 7: 10.
11. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, et al. (2004) 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* 4: 38.
12. Oxman AD, Lavis JN, Lewin S, Fretheim A (2009) SUPPORT tools for evidence-informed health policymaking (STP) 1: What is evidence-informed policymaking? *Health Res Policy Syst* 7 Suppl 1: S1.
13. Thomas J, Harden A (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 8: 45.
14. Spencer L, Ritchie J, Lewis J, Dillon L (2003) Quality in qualitative evaluation: a framework for assessing research evidence. England: Government Chief Social Researcher's Office.
15. Pawson R (2002) Evidence-based policy: the promise of 'realist' synthesis. *Evaluation* 8: 340–358.
16. Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, et al. (2004) Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technol Assess* 8: 1–IV.
17. Kiwanuka SN, Kinengyere AA, Rutebemberwa E, Nalwadda C, Sengooba F, et al. (2011) Dual practice regulatory mechanisms in the health sector: a systematic review of approaches and implementation. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
18. Carroll C, Booth A, Cooper K (2011) A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Med Res Methodol* 11: 29.
19. Pawson R (2006) Digging for nuggets: how 'bad' research can yield 'good' evidence. *Int J Soc Res Methodol* 9: 127–142.
20. Jefferson T, Demicheli V, Vale L (2002) Quality of systematic reviews of economic evaluations in health care. *JAMA* 287: 2809–2812.
21. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, et al. (2008) Incorporating considerations of resources use into grading recommendations. *BMJ* 336: 1170–1173.
22. Treweek S, Zwarenstein M (2009) Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 10: 37.
23. Lewin S, Lavis JN, Oxman AD, Bastias G, Chopra M, et al. (2008) Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews. *Lancet* 372: 928–939.
24. Chopra M, Munro S, Lavis JN, Vist G, Bennett S (2008) Effects of policy options for human resources for health: an analysis of systematic reviews. *Lancet* 371: 668–674.
25. Donner A, Klar N (2004) Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 94: 416–422.
26. Varnell SP, Murray DM, Janega JB, Blitstein JL (2004) Design and analysis of group-randomized trials: a review of recent practices. *Am J Public Health* 94: 393–399.

27. WHO (2010) Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations. Geneva: World Health Organization.
28. Huicho L, Dieleman M, Campbell J, Codjia L, Balabanova D, et al. (2010) Increasing access to health workers in underserved areas: a conceptual framework for measuring results. *Bull World Health Organ* 88: 357–363.
29. Glasziou P, Chalmers I, Altman DG, Bastian H, Boutron I, et al. (2010) Taking healthcare interventions from trial to practice. *BMJ* 341: c3852.
30. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, et al. (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336: 924–926.
31. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A (2011) GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol* 64: 380–382.
32. Peters DH, El-Saharty S, Siadat B, Janovsky K, Vujicic M (2009) Improving health service delivery in developing countries: from evidence to action. Washington (D.C.): World Bank.
33. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, et al. (2009) SUPPORT tools for evidence-informed health policymaking (STP) 9: assessing the applicability of the findings of a systematic review. *Health Res Policy Syst* 7 Suppl 1: S9.
34. Wang S, Moss JR, Hiller JE (2006) Applicability and transferability of interventions in evidence-based public health. *Health Promot Int* 21: 76–83.
35. Victora CG (2010) Commentary: LiST: using epidemiology to guide child survival policymaking and programming. *Int J Epidemiol* 39 Suppl 1: i1–i2.
36. Bryce J, Friberg IK, Kraushaar D, Nsona H, Afenyadu GY, et al. (2010) LiST as a catalyst in program planning: experiences from Burkina Faso, Ghana and Malawi. *Int J Epidemiol* 39 Suppl 1: i40–i47.
37. Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, et al. (2011) Evidence summaries tailored to health policy-makers in low- and middle-income countries. *Bull World Health Organ* 89: 54–61.
38. Rosenbaum SE, Glenton C, Nylund HK, Oxman AD (2010) User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. *J Clin Epidemiol* 63: 607–619.
39. Bosch-Capblanch X and Project Team (2011) Handbook for developing health systems guidance: supporting informed judgements for health systems policies. Basel: Swiss Tropical and Public Health Institute, Available: <http://www.swisstph.ch/?id=53>; http://www.swisstph.ch/fileadmin/user_upload/Pdfs/SCIH/WHOHS_G_Handbook_v04.pdf. Accessed 10 February 2012.
40. Bennett S, Agyepong IA, Sheikh K, Hanson K, Ssenooba F, et al. (2011) Building the field of health policy and systems research: an agenda for action. *PLoS Med* 8: e1001081. doi:10.1371/journal.pmed.1001081.
41. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, et al. (2011) GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64: 383–394.
42. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, et al. (2010) AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol* 63: 1308–1311.
43. Kane SS, Gerretsen B, Scherpier R, Dal PM, Dieleman M (2010) A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Serv Res* 10: 286.
44. Glenton C, Lewin S, Scheel IB (2011) Still too little qualitative research to shed light on results from reviews of effectiveness trials: a case study of a Cochrane review on the use of lay health workers. *Implement Sci* 6: 53.
45. Admassie A, Abebaw D, Woldemichael AD (2009) Impact evaluation of the Ethiopian Health Services Extension Programme. *J Dev Effect* 1: 430–449.
46. Oxford Policy Management (2009) Lady health worker programme. External evaluation of the National Programme for Family Planning and Primary Health Care. Management review. Oxford: Oxford Policy Management.
47. Lewin S, Munabi-Babigumira S, Glenton C, Daniels K, Bosch-Capblanch X, et al. (2010) Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases. *Cochrane Database Syst Rev* CD004015.
48. Lewin S, Oxman AD, Glenton C (2006) Assessing healthcare interventions along the complex-simple continuum: a proposal [P100]. *Cochrane Colloquium* 2006.
49. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337: a1655.