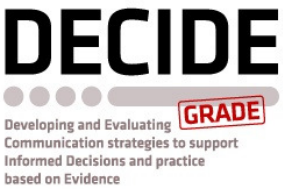


Project no. 258583
Project acronym: DECIDE
Project title: Developing and Evaluating Communication strategies to support Informed Decisions and practice based on Evidence

SP1 - Cooperation

Collaborative Project
Small or medium scale research project

FP7 – HEALTH – 2010 – two-stage

Deliverable 5.1 – Strategies for effectively and efficiently communicating evidence-based health systems recommendations

Due date of deliverable: 31 December 2013
Actual submission date: 20 December 2013

Start date of project: 01 January 2011 Duration: 60 Months

Organisation name of lead beneficiary for this deliverable: Nasjonalt Kunnskapssenter for Helsetjenesten

Coordinator: Professor Shaun Treweek

Signed: S. Treweek Date: 20/12/2013

Deliverable 5.1 – Strategies for effectively and efficiently communicating evidence-based health systems recommendations

Nature: Report

Dissemination Level: Public (PU)

Owner

Name:	KS
Lead Beneficiary:	KS
E-mail:	oxman@online.no

Context

Author(s):	KS
Work Package:	WP5

Document Status

Version:	1.0
Last modified:	Friday 20th December 2013
Status:	Released
Approved by:	Shaun Treweek and Andy Oxman
Date Approved:	20 December 2013

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Appendices

Background

In order to maximize use of available resources and make the right health care decisions, health policymakers and managers, in addition to clinicians, patients and the public, need reliable up-to-date evidence about "what works".[1] Systematic reviews are valuable sources of research evidence for informing policy decisions. They are based on a comprehensive search for, and appraisal of, relevant studies, so the chances of being misled are greatly diminished compared to relying on a single study or a non-systematic review.[1,2,3,4] Less time and skill is needed to find and appraise the evidence. In addition, a systematic review illuminates the areas where evidence is lacking and further evaluation is needed.[1,5,6,7] However these reviews are usually written for scientific or clinical audiences and are not necessarily well tailored for the information needs of policymakers.[8]

A number of review-derived products for policymakers now exist, including summaries of systematic reviews, overviews of systematic reviews and policy briefs.[9,10] Collections of such summaries are also becoming available, such as SUPPORT Summaries,[11] Rx for Change,[12] and Cochrane Reviews Identified and Summarised for Policy (CRISP).[13]
One of the most important predictors of policymakers' use of systematic reviews is that they are easy to use.[14] However, there are few reported evaluations of review or summary formats for policymakers. What studies there are support the use of a graded-entry format and up-front take-home messages for HTA reports,[2,15] the use of language that is tailored to non-clinical audiences,[16] and inclusion of content that helps users to contextualize the evidence (discussion of applicability and relevance).[2,15,16] These factors are likely to also be applicable for the delivery of information from reviews to patients and the public.

The FP6-funded SUPPORT Collaboration (http://www.support-collaboration.org/) used the GRADE system to prepare summaries of systematic reviews for policymakers and managers in low and middle-income countries.[11,15,17] We are not aware of similar studies in high-income countries. One of the overriding findings of user testing of the SUPPORT summaries was the desire for clear and short messages. Other studies where policymakers have been interviewed for their

preferences regarding presentation of research have had similar findings.[2,18,19] There is, however, a limit to how short informative messages can be before they lose their scientific value or credibility. When these limits are reached, devices other than text editing must be used, such as graded entry structuring of the text with a first page summary of key messages.

Poor comprehension of numerical information and statistics may also be a problem. Studies show that even highly educated people struggle to understand risks.[20] Appreciation of the value (and limitations) of systematically reviewed evidence is dependent on a basic understanding of the underlying concepts. People with limited exposure to research may not have developed correct conceptual models of this kind of information. This can result in frustration over unmet expectations or poor understanding of messages. However, correct comprehension is not only dependent on the skills and knowledge of the reader, but also on the characteristics of the information.[21] By anticipating weak background knowledge or low levels of statistical literacy, extra information can be provided to help readers better understand the strengths and limitations of scientific evidence. Provision of explicit meta-information (such as an explanation of what is meant by “certainty of the evidence”) may help replace frustration with reflection, for instance in the case of weak or missing evidence. This problem is shared with other DECIDE work packages, in particular WPs 2 and 3, and is also addressed in their ongoing work.

We have not found any published evaluations of strategies for disseminating health technology assessments or recommendations to policymakers and managers. As noted above, recommendations and decisions depend on information and judgements that are beyond the scope of systematic reviews. DECIDE will help to fill this gap in knowledge. Whether or not recommendations are made, the challenge of structuring a policy problem can be supported by introducing some of the main elements that need to be taken into further consideration in specific settings. This includes structured consideration of the applicability of the evidence, impacts on equity, and costs. Even when specific answers are not available, policymakers find this information useful. When the evidence is too uncertain to provide clear answers, or

decision makers' settings vary greatly from those in the studies, this may be the best way of rendering knowledge from research useful for policy decisions.

We have addressed these problems by building on previous work, and developing and evaluating three strategies to communicate evidence-based health system and population health recommendations effectively and efficiently:

- interactive Summary of Findings to facilitate understanding and use of the results of systematic reviews in health system and population health recommendations and decisions
- interactive Evidence to Decision frameworks to facilitate going from evidence to health system and population health recommendations and decisions
- explanations of terms relevant to health system and population health recommendations and decisions

Methods

Our objective was to develop and evaluate strategies to communicate evidence-based health system and population health recommendations effectively and efficiently. The strategies were identified, developed and evaluated iteratively using brain storming, stakeholder feedback, a survey, user testing with health policymakers and managers, and testing the strategies with actual guidelines. Further development and testing of the three strategies that we have developed is ongoing.

Strategy development and user testing

Brainstorming workshops

We had a series of brainstorming sessions: first to generate potential strategies; then to generate ideas about how to develop those strategies; and then iteratively to generate solutions to problems identified in the strategies through feedback, user testing and experience of using the strategies. At the first brainstorming workshop, we used examples of SUPPORT Summaries for health system policies and recommendations developed using the GRADE approach.[11,15,17] We identified problems with these examples and ways in which dissemination could be improved by changes in presentation, accompanying materials, and supportive strategies.

Members of the research team in Oslo, other DECIDE partners and other members of the GRADE Working Group participated in these discussions.

Stakeholder group feedback

We established an advisory group with 32 people from 15 countries, including policy makers, researchers, journalists, health economists, clinicians and consumer advocates. Members of the advisory group were identified by all of the DECIDE partners. We sought feedback from the advisory group by email on the strategies we developed.

Survey of policymakers

In parallel with the first consultation of the advisory group, we conducted an online survey of a diverse group of stakeholders with experience making health system decisions.[22] We asked about perceptions of criteria relevant to making health system decisions, use of evidence, grading systems, and evidence summaries. Details of the methods used in the survey are described in the published report of this survey (Appendix 1).[23]

User testing

We conducted user tests of the strategies with policymakers and others in our target audiences. We followed a methodology used for similar work by Partner 2 (the Norwegian Knowledge Centre for the Health Services).[15,24,25,26,27,28] The user tests were performed individually and took approximately one hour. With the participant's written permission, we audio and video-recorded each test, and an observer took notes. Using a semi-structured interview guide, we considered both immediate first impressions and detailed exploration. The interview guide was designed to explore six of the seven different facets of "user experience" as described in a model by Peter Morville: usability (defined for our purposes as "correct understanding and ease of use"), credibility, usefulness, desirability, findability and value. The seventh facet from this model – accessibility – was not addressed, as online accessibility was not relevant. Follow-up questions covered overall impressions and suggestions for improvement. Transcriptions were made of the recorded interviews.

We reviewed all of the notes and transcriptions, looking primarily for barriers and facilitators related to correct interpretation, ease of use and favourable reception. We traced findings back to specific elements or characteristics of the strategies that appeared to facilitate or hinder problems. We rated findings in three categories according to the severity of the problem for the user: high (causes incorrect interpretation, critical errors or high degree of uncertainty or dissatisfaction), medium (causes much frustration or unnecessarily slow use), low (minor or cosmetic problems). We also registered nice-to-haves (things users explicitly liked) or suggestions for improvement.

Testing our strategies with real guidelines

We have tested the strategies by using them in real guideline processes, including WHO guidelines and public health guidelines. Use of the strategies by other organisations making health system or public health recommendations or decisions is ongoing. We have collected structured feedback from organisations using the strategies and from workshops that we have organised for organisations responsible for health system or public health recommendations or decisions.

Evaluating the strategies

We will evaluate the strategies in randomized controlled trials during the last two years of the project. The objectives of the trials are to assess the impacts of the strategies or specific elements of the strategies on correct comprehension of key information, attitudes, hypothetical or intended behaviour, and satisfaction. We will interview selected participants to explore strengths and weaknesses of the strategies, reasons for their success or failure, and ways of improving them. We will also continue with further iterations of user testing, testing the strategies with real guidelines and decisions, feedback from our advisory group and other users, and brainstorming to generate ideas for addressing problems and improving the strategies.

Results

Interactive Summary of Findings

It can be difficult for doctors, patients, or policy makers to find reliable research results about the effect of treatments or other interventions when making decisions about health care. Systematic reviews – based on a systematic collection, appraisal and analysis of primary studies – can save people both time and work. But due to their complexity and length, the most important messages from a review can be difficult both to locate within the document and to understand. Summary of Findings (SoF) tables have been developed to alleviate some of these problems, with an aim to make evidence from systematic reviews more accessible to a wide audience.[26,27]

The SoF table presents the key messages from a systematic review in a concise format. The table is an output from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, a structured transparent system for systematic review authors to guide them through the process of selecting the most patient-important outcomes, reporting the evidence for these and rating the certainty of that evidence.[29] SoF tables include seven elements that have been judged to be most critical when making a health care decision (see Box 1). These judgments are the cumulative result of efforts over the last decade of the GRADE working Group and the Cochrane Applicability and Recommendations Methods Group.[30]

Box 1. Seven elements of a Summary of Findings table

1. A list of the most important outcomes, both desirable and undesirable
2. A measure of the typical burden of these outcomes (e.g. control group, estimated risk)
3. A measure of the risk in the intervention group or, alternatively or in addition, a measure of the difference between the risks with and without intervention
4. The relative magnitude of effect
5. Numbers of participants and studies addressing these outcomes
6. A rating of the overall confidence in the effect estimate for each outcome (which may vary by outcome); and possibly
7. Comments

We received 112 responses (70% response rate) to our survey of stakeholders. Most respondents had healthcare (85%) and research (79%) experience. They (99%) supported the use of evidence from other countries (94%) and grading systems (81%). Regarding the SoF table, respondents had divergent views regarding whether the same (38%) or different (45%) grading systems should be used across different

types of health decisions. Over 90% of respondents rated all components of our evidence summary as important.

We have improved the SoF format by making it both simpler and more comprehensive, as well as more flexible. Our goal is to make the table compatible with the needs of a wide range of users, as well as different types of data and use contexts. To achieve this, we designed an electronic, interactive Summary of Findings table (iSoF) that allows table producers to tailor the presentation to their target audience by adjusting which outcomes and how much information about those outcomes are displayed. More importantly, the new solution will allow users to interact with the table by adding or hiding outcomes, by adding or hiding information about those outcomes, by viewing results as numbers, text or graphic representations, and by accessing explanations of standard concepts (such as confidence intervals) and topic specific explanations provided by the producer. For a more detailed list of specifications for development, see Box 2.

Box 2 New iSoF features

Features for users:

- Simple, user-friendly interface
- Layered presentation of information, allowing users to initially view a simple table with a minimum amount of information and (if desirable) drill down to more details, including links to reviews and full evidence profiles
- Providing users with control over their viewing choices, including which outcomes to show in detail and how to view the results for these outcomes (as text, numbers or graphic representations)
- Providing step-by-step visual presentation of the absolute effects and absolute differences, that includes an explanation of the confidence intervals in a way that makes them easy to grasp and see why they are important
- Providing interactive explanations of generic terms (replacing legends and glossaries)
- Providing interactive explanations of table-specific terms (replacing footnotes)
- Responsive formatting for use on different size screens/devices
- Availability in different languages

Features for producers:

- Template flexibility that can accommodate data from different kinds of reviews, including those without meta-analysis
- Ability to enter (and present) different levels of baseline (control group) risk for each outcome
- Control over which information is expanded/displayed (and which is collapsed/hidden) in the initial (default) presentation, including:
  - Which outcomes
  - What information about each outcome
  - Which baseline risk (including more than one for outcomes when this is relevant)
- Automatic reminders to include some information that is essential for understanding the findings of a systematic review, but is sometimes missing, including explanations about scales, about where the estimates of baseline risk came from, and about the reasons for downgrading or upgrading the certainty of evidence.
- Allowing producers to tailor their own template, for instance to rearrange the order of the columns, create a custom default presentation, or add organization logos
- Templates for table production in different languages

Based on feedback and user testing we have revised the iSoF. Figure 1 is a screenshot of an iSoF for a health systems decision, showing all of the information for each outcome.

Figure 1. iSoF showing all columns

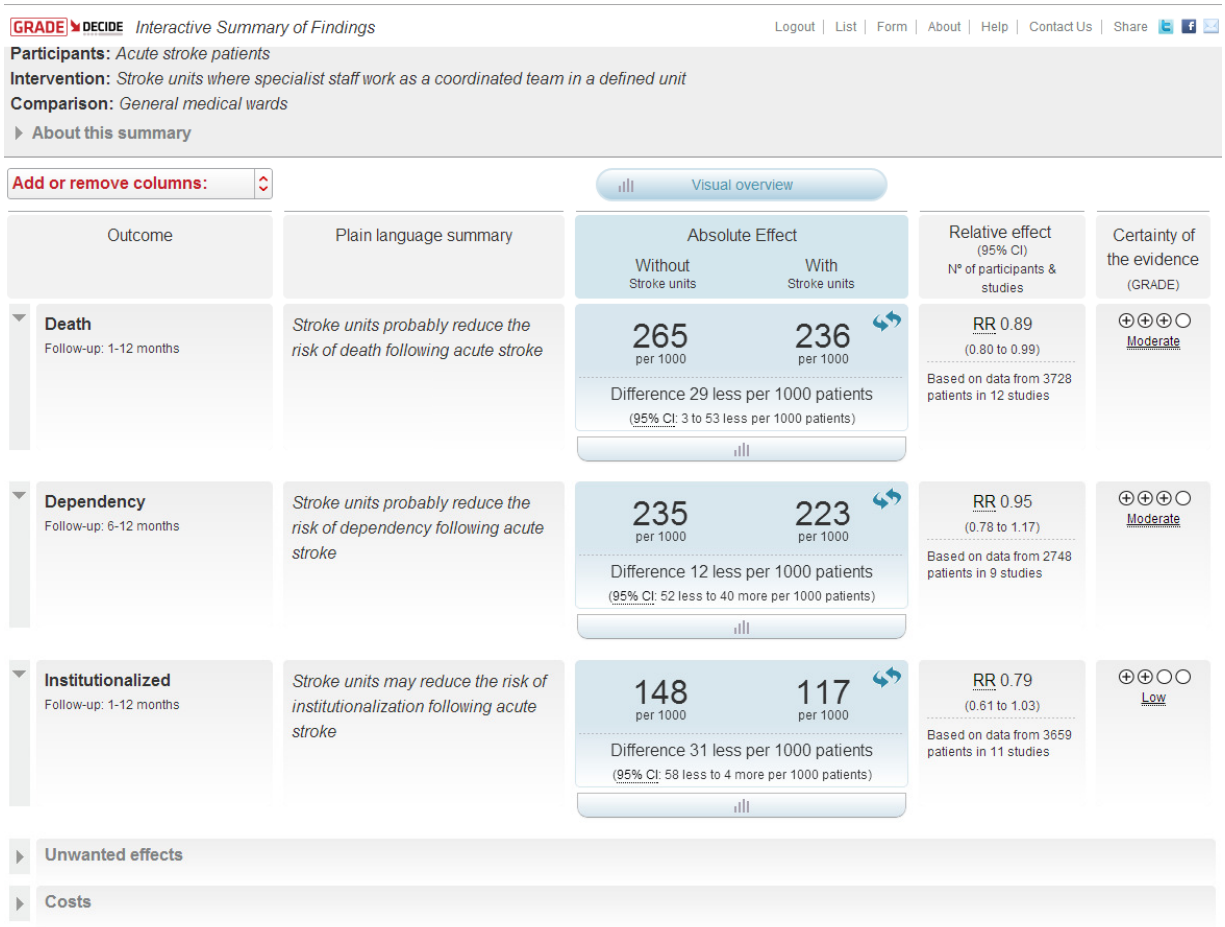


Figure 2 is a screen shot of a simplified table.

Figure 2. iSoF showing selected columns

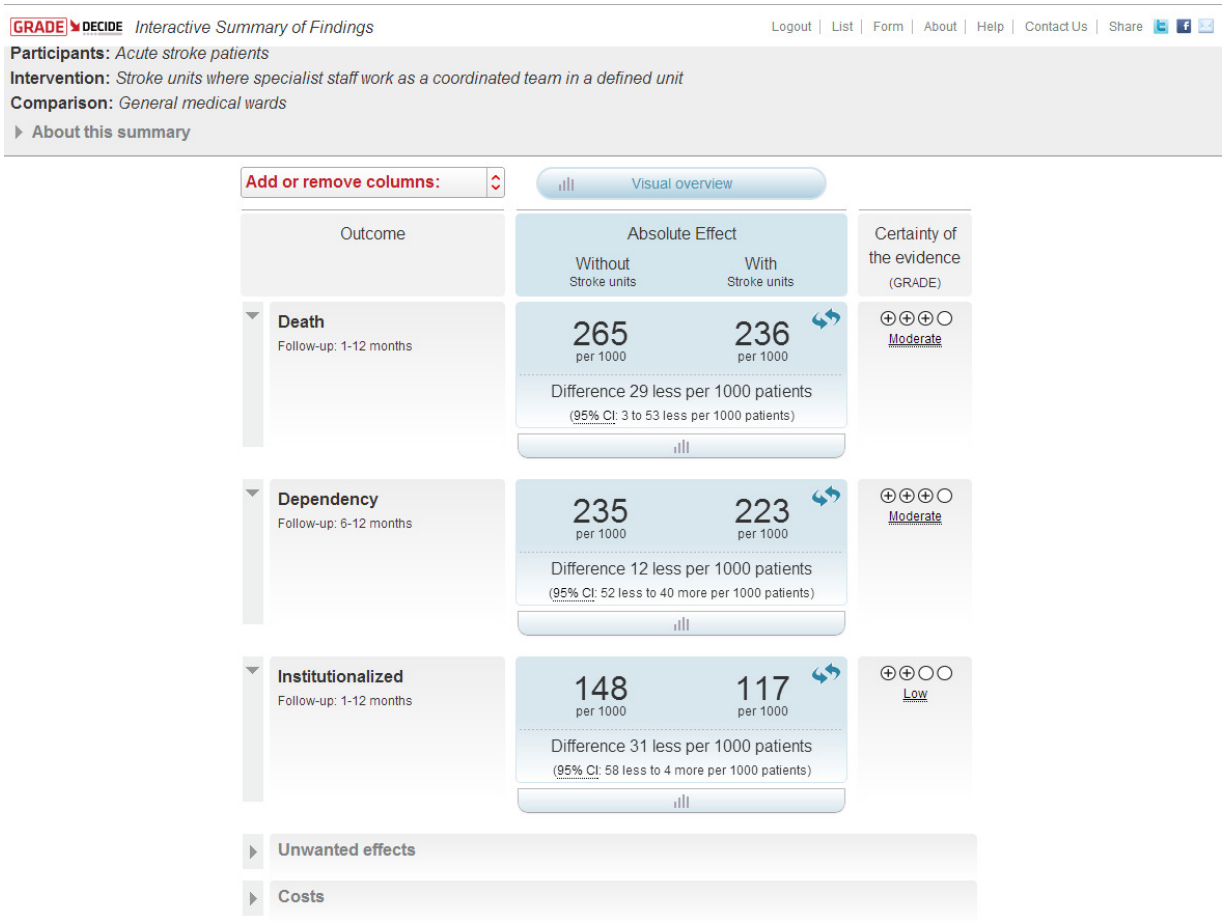


Figure 3 is a screenshot of the text that can be viewed by flipping one of the cells showing the absolute effect.

Figure 3. iSoF with the absolute effect in text

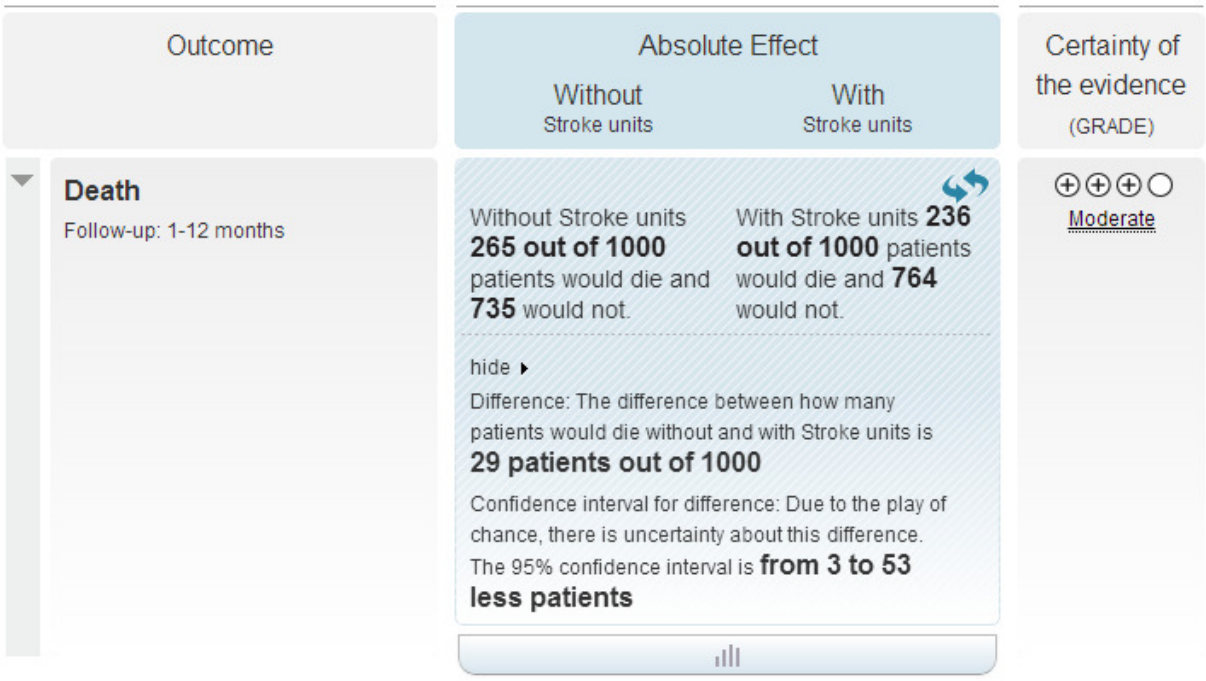


Figure 4 is a screenshot from the visualisation for one of the outcomes.

Figure 4. iSoF visualisation of an outcome

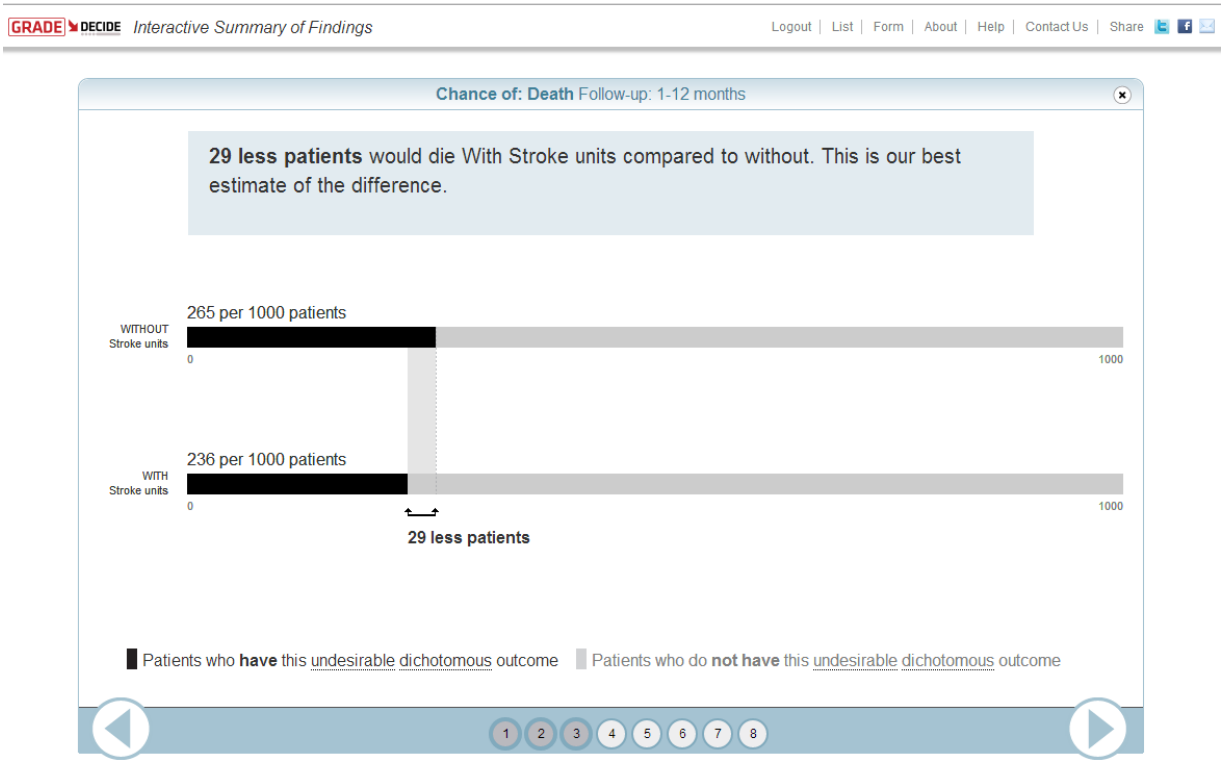
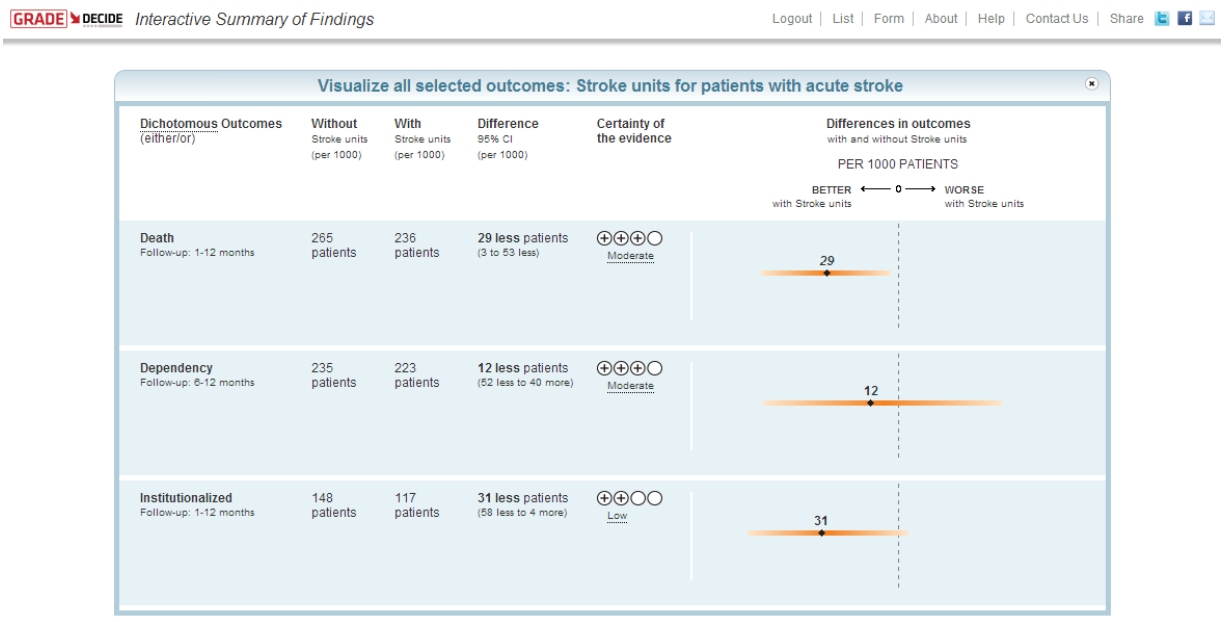


Figure 5 is a screenshot of the visualisation for all selected outcomes.

Figure 5. iSoF visualisation of all selected outcomes



Interactive Evidence to Decision Frameworks

The evidence to decision framework is a systematic and transparent approach for going from evidence to recommendations, coverage decisions or health system

decisions. It includes criteria, judgements that must be made in relation to each criterion, evidence to inform each judgement and conclusions based on an overall judgement across all of the criteria. It helps to ensure that important factors that determine a decision (the criteria) are considered, structure discussion, identify reasons for disagreements and make the basis for a decision transparent to those affected.

The initial evidence to decision framework included ten criteria:

1. Seriousness of the problem
2. Number of people affected
3. Quality of evidence
4. Benefits/desirable) effects
5. Adverse (undesirable) effects
6. Resource use (cost)
7. Value for money
8. Impacts on equity
9. Implementability (feasibility)
10. Acceptability

Respondents to our survey of stakeholders (99%) indicated that systematic consideration of the available evidence would help to improve health system decision-making processes. They rated all ten criteria in the DECIDE framework (listed above) as important in the decision-making process.

Most of our advisory group indicated that

- There were not important considerations that were missing (60%)
- There were not considerations included in the framework that should not have been (87%)
- The framework was applicable to different types of health system decisions (67%)
- It was not more complicated than necessary (80%)
- It was organised in a logical way (87%)
- The criteria were clearly labelled and explained (73%)

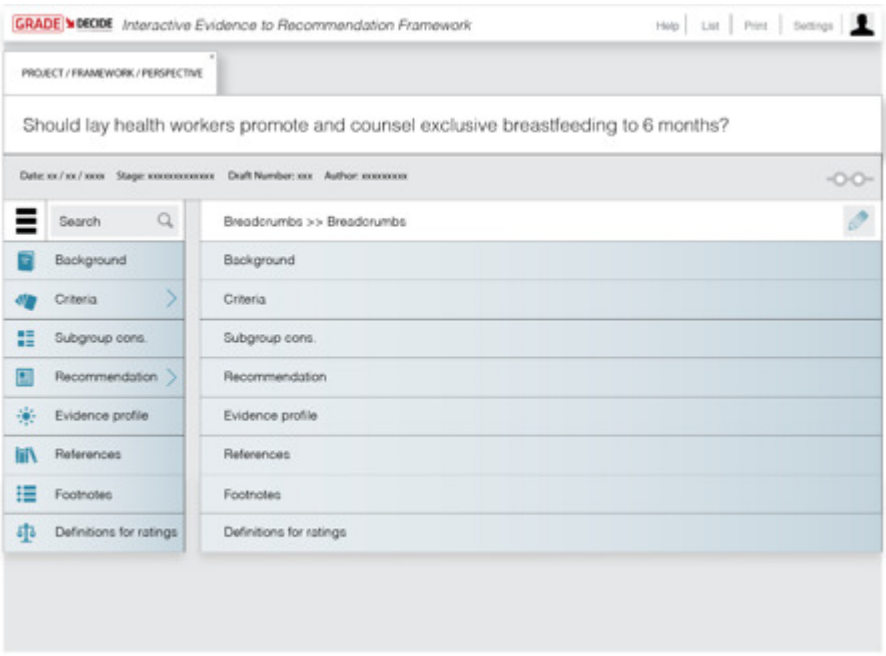
- The framework was suitable for informing and helping people to make health system decisions (73%)
- It was likely to be useful to people responsible for health system decisions (60%)
- Overall it was adequate (53%) or partially adequate (40%) for its intended purpose

Their comments and our responses are summarised in Appendix 2.
Based on feedback from the advisory group, feedback from guidelines panels that used the framework and user testing, we made a number of changes to the framework. In parallel we participated in a series of discussions across work packages 1 (recommendations for health professionals), 2 (coverage decisions), 3 (recommendations for patients), 4 (recommendations for diagnostic tests), 5 (health system decisions) and 6 (strategies for collaboration) with the aim of harmonising evidence to decision frameworks for different types of decisions and audiences. This resulted in the development of a generic framework that can be tailored to different audiences, different types of decisions and different contexts.

The generic evidence to decision framework is attached as Appendix 3, explanations as Appendix 4 and an example of a framework for a health system decision as Appendix 5.

The interactive Evidence to Decision framework (Figure 6) will enable organisations to tailor the framework, reports and interactive resources to help target audiences to go from evidence to a decision. Organisations will be able to modify the terminology, explanations, criteria and response options. They will be able to generate tailored interim reports (e.g. to consult decision makers or stakeholders) and final reports (e.g. tables or appendices to a recommendation or decision). They will also be able to generate interactive resources to assist decision makers to go from evidence (with or without a recommendation) to a decision.

Figure 6. Interactive Evidence to Decision Framework



Explanations of terms

We have developed a glossary in collaboration with work package 3 (for patients and the public) and work package 2 (for policymakers). Well-informed choices about how to intervene to improve health outcomes depend on access to good information, particularly research evidence. The use of jargon can be a barrier to people's understanding and use of research evidence to inform their choices. Inconsistent use of language also can cause confusion. The aim of this glossary is to facilitate informed healthcare choices by promoting consistent use of plain language and providing plain language explanations of concepts and terms that people might need to understand in order to assess claims about treatments. This includes claims arising from summaries of research evidence (systematic reviews) and evidence-informed recommendations.

The glossary is an open access database that can facilitate the understanding and use of a variety of resources, including:

- plain language summaries
- summaries of findings
- evidence to decision frameworks
- databases of systematic reviews, recommendations or patient information

The glossary includes:

- brief plain language definitions (that can be used as scroll over explanations)
- longer explanations
- links to resources such as illustrative examples, videos or interactive applications that help people to understand or apply the term or concept
- synonyms
- suggested plain language terms
- technical definitions

The glossary can be used by guideline producers, health technology assessment agencies and others providing support for evidence-informed healthcare decisions, including health system decisions. Among other uses, it provides explanations for terms used in interactive Summaries of Findings and interactive Evidence to Decision frameworks. Organisations can utilise some or all of those features and some or all of the terms that are included in the glossary. The glossary currently includes over 120 terms.

Conclusions

Deliverable 5.1 is complete and we have met our objectives up to this phase of the project. Work is ongoing and the three components of this deliverable are at different stages. We have completed user testing of the first version of the interactive Summary of Findings. The next version will have a responsive design, so that it can be viewed on different screens, as well as improvements based on the user testing. We have completed user testing of the paper version of the Evidence to Decision (EtD) framework. A fully functional interactive version will be available by June, in time for the conference that we are organising in Edinburgh. We will complete user testing of the interactive EtD framework and the glossary next year. Testing the three strategies in actual guidelines and decisions is ongoing. Trials are planned for years four and five of the project (2014 – 2015).

.

References

[1] Lavis J, Posada F, Haines A, Osei E. Use of research to inform public policymaking. The Lancet 2004; 364:1615.

[2] Lavis JN, Davies HTO, Oxman A, Denis JL, Golden-Biddle K, Ferlie E: Towards systematic reviews that inform healthcare management and policymaking. J Health Serv Res Policy 2005, 10:35-48.

[3] Oxman AD, Lavis JN, Lewin S, Fretheim A: SUPPORT Tools for evidence-informed health Policymaking (STP). 1. What is evidence-informed policymaking. Health Res Policy Syst. 2009, 7(Suppl 1):S1.

[4] Lavis JN, Oxman AD, Grimshaw J, Johansen M, Boyko JA, Lewin S, Fretheim A: SUPPORT Tools for evidence-informed health Policymaking (STP). 7. Finding systematic reviews. Health Res Policy Syst. 2009, 7(Suppl 1):S7.

[5] Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. Ann Intern Med 1997; 127:37-42.

[6] Nutley SM, Davies HTO, Smith PC. What works?: evidence-based policy and practice in public services: The Policy Press; 2000.

[7] Lavis J, Davies HTO, Gruen R, Walshe K, Farquhar C. working within and beyond the Cochrane Collaboration to make systematic reviews more useful to healthcare managers and policymakers. Healthcare Policy. 2006;1:21-33.

[8] Sheldon T. Making evidence synthesis more useful for management and policy-making. J Health Serv Res Policy. 2005;10(Suppl 1):1 - 5.

[9] Lavis JN, Permanand G, Oxman AD, Lewin S, Fretheim A: SUPPORT Tools for evidence-informed health Policymaking (STP). 13. Preparing and using policy briefs to support evidence-informed policymaking. Health Res Policy Syst. 2009, 7(Suppl 1):S13.

[10] Lavis J. Supporting the Use of Systematic Reviews in Policymaking. PLoS Med. 2009; 6(11):
e1000141. doi:10.1371/journal.pmed.1000141

[11] SUPPORT Summaries. Evidence of the effects of health system interventions for low- and middle-income countries. Available at: www.supportsummaries.org/

[12] Canadian Agency for Drugs and Technologies in Health. Rx for Change. Available at: www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions/

[13] EvidenceMap.org. Cochrane Reviews Identified and Summarised for Policy (CRISP). Available at: www.policymakers.evidencemap.net/

[14] Dobbins M, Cockerill R, Barnsley J. Factors affecting the utilization of systematic reviews. Int J Technol Assess Health Care 2001; 17:203-14.

[15] Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, Young T, Althabe F, Ciapponi A, Garcia Marti S, Meng Q, Wang J, De la Hoz Bradford AM, Kiwanuka S, Rutebemberwa E, Flottorp S, Oxman AD. Evidence summaries tailored for health

policymakers in low and middle-income countries. WHO Bull 2011; 89:54-61.

[16] Turner T, Green S, Harris C. Supporting evidence-based health care in crises: What information do humanitarian organisations need? Disaster Medicine and Public Health Preparedness. In press.

[17] SUPPORT Collaboration. SUPPORT structured summaries of systematic reviews. http://www.support-collaboration.org/summaries.htm

[18] Dobbins M, Jack S, Thomas H, Kothari A. Public health decision-makers' informational needs and preferences for receiving research evidence. Worldviews on Evidence-Based Nursing 2007; 4:156-63.

[19] Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1. The reality according to policymakers. J Epidemiol Community Health. 2004; 58:811-6.

[20] Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001; 21:37-44.

[21] Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc. 2007; 14:713-21.

[22] Vogel JP, Oxman AD, Glenton C, Rosenbaum S, Lewin S, Gülmezoglu AM, Souza JP. Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: An international survey. Health Research Policy and Systems 2013; 11:19 DOI: 10.1186/1478-4505-11-19.

[23] Vogel JP, Oxman AD, Glenton C, Rosenbaum S, Lewin S, Gülmezoglu AM, Souza JP. Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: An international survey. Health Research Policy and Systems 2013; 11:19 DOI: 10.1186/1478-4505-11-19.

[24] Glenton C, Underland V, Kho M, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. J Clinical Epidemiol 2006; 59:770-8.

[25] Rosenbaum SE, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library. BMC Medical Informatics and Decision Making. 2008;8(1):34.

[26] Rosenbaum SE, Glenton C, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane Reviews. J Clin Epidemiol 2010; 63:607-19.

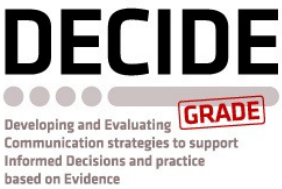
[27] Rosenbaum SE, Glenton C, Oxman AD. Summary of Findings tables improved understanding and rapid retrieval of key information in Cochrane Reviews. J Clin Epidemiol 2010; 63:620-6.

[28] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the Results of Cochrane Systematic Reviews to a Consumer Audience: A Qualitative Study. Med Decis Making. 2010;30(5):566-77.

[29] Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines 1.

Introduction - GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011; 64:383-94.

[30] Langendam M, Akl E, Dahm P, Glasziou P, Guyatt G, Schünemann H. Assessing and presenting summaries of evidence in Cochrane Reviews. Submitted for publication.



Appendix 1

Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: An international survey

HEALTH RESEARCH POLICY AND SYSTEMS

RESEARCH Open Access

Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey

Joshua P Vogel[1,2], Andrew D Oxman[3*], Claire Glenton[3], Sarah Rosenbaum[3], Simon Lewin[3,4], A Metin Gülmezoglu[2] and João Paulo Souza[2]

Abstract

Background: The DECIDE framework was developed to support evidence-informed health system decisions through evidence summaries tailored to health policymakers. The objective of this study was to determine policymakers' perceptions regarding the criteria in the DECIDE framework and how best to summarise and present evidence to support health system decisions.

Methods: We conducted an online survey of a diverse group of stakeholders with health system decision experience from 15 countries and the World Health Organization. We asked about perceptions of criteria relevant to making health system decisions, use of evidence, grading systems, and evidence summaries.

Results: We received 112 responses (70% response rate). Most respondents had healthcare (85%) and research (79%) experience. They (99%) indicated that systematic consideration of the available evidence would help to improve health system decision-making processes and supported the use of evidence from other countries (94%) and grading systems (81%). All ten criteria in the DECIDE framework were rated as important in the decision-making process. Respondents had divergent views regarding whether the same (38%) or different (45%) grading systems should be used across different types of health decisions. All components of our evidence summary were rated as important by over 90% of respondents.

Conclusions: Survey respondents were supportive of the DECIDE framework for health system decisions and the use of succinct summaries of the estimated size of effects and the quality of evidence. It is uncertain whether the findings of this survey represent the views of policymakers with little or no healthcare and research experience.

Keywords: Decide, Evidence summaries, Health system decisions

Background

Decisions regarding health systems are often political processes involving a number of policymakers and other key stakeholders. There is a need to support these decisions with the best available evidence, however, the effects of health system interventions are often uncertain and stakeholders may have a low level of medical and research literacy. The Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE) project is a collaborative research project funded by the European Commission's Seventh Framework Programme [www.decide-collaboration.eu]. The project's objective is to develop and evaluate communication strategies to support evidence-informed decisions by building on the work of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group [www.gradeworkinggroup.org] and the Cochrane Applicability and Recommendations Methods Group [www.armg.cochrane.org]. As part of the DECIDE project, we have developed a framework

* Correspondence: oxman@online.no
[3]Global Health Unit, Norwegian Knowledge Centre for the Health Services, P O Box 7004, St. Olavsplass, 0130, Oslo, Norway
Full list of author information is available at the end of the article

for communicating evidence to inform decisions about health systems [http://www.decide-collaboration.eu/WP5/Strategies/Framework].

The development of this framework was informed by our earlier work on plain language summaries of systematic reviews [1] and work on evidence-based policy briefs [2-4]. The framework includes relevant criteria for making health system decisions as well as evidence to inform judgements about each criterion. The criteria included in the framework are seriousness of the problem, number of people affected, quality of the evidence, size of the benefits, size of the adverse effects, resource use (costs), value for money, impacts on equity, implementability (feasibility), and acceptability. These framework criteria emerged from a literature review, brainstorming, feedback from stakeholders, and application of the framework to examples. This survey, along with user testing and further application of the framework to health system decisions, will further inform the selection of criteria included in the framework and how they are presented.

There are several different systems available to grade the quality of evidence on the effects of healthcare interventions. Most of these have been used primarily for clinical practice guidelines and those systems have become increasingly similar to GRADE or replaced by GRADE [5-7]. Other systems have been used for population health guidelines (primarily for public health rather than health system interventions), such as the systems used by the Task Force on Community Preventive Services [6,8,9], although the GRADE system is also used for public health and health system recommendations [10,11]. All of these grading systems have focused primarily on grading the quality of evidence and only to a lesser extent, if at all, on frameworks for going from evidence to public health or health system decisions or recommendations.

The use of grading systems for rating the quality of evidence is seen as essential for evidence-based guideline development. The GRADE methodology has been adopted by WHO, as well as by national guideline developers and others to assess the quality of evidence in guideline development as it enables a comprehensive, transparent and structured analysis of available literature, whilst clearly communicating the quality of evidence and strength of the recommendations [1,12,13]. However, there is considerable debate as to whether the same approach to grading evidence for clinical decisions should also be used for health system decisions. Advocates for the use of the same grading system for both types of decisions state it can minimise confusion, reduce the risk of bias, and maintain transparency and consistency across different types of decisions. Critics state that a rating system that is appropriate for clinical interventions may not discriminate evidence in a way that is appropriate for health system programmes or policies, potentially disadvantaging effective interventions that are not amenable to randomized controlled trials [10].

The presentation of evidence to inform health system decisions requires the development of summaries that can 1) communicate complex information in plain language; 2) clearly present the anticipated effects and the certainty (quality) of the underlying evidence; and 3) effectively convey uncertainty. We have previously developed and user-tested evidence summaries tailored for health policymakers in low- and middle-income countries [2], built on earlier findings for Cochrane reviews [14,15]. An evidence summary based on the DECIDE framework needs to clearly communicate evidence relating to each of the ten DECIDE framework criteria, including the effect sizes and quality of evidence available for the anticipated desirable and undesirable effects of the intervention being considered. The design of these evidence summaries to support policymakers' understanding of evidence therefore requires a detailed analysis of their perceptions of certainty (quality) of the evidence and how it is assessed.

To obtain a better understanding of policymakers', managers' and other stakeholders' perceptions of the criteria for health system decisions in the DECIDE framework, and how evidence on these should be presented, we conducted an international online survey. The survey focused on stakeholders' perceptions of each of the ten DECIDE framework criteria, the grading of health system evidence, and the use of evidence summaries. This paper aimed to collect information regarding the experience and perceptions of participants with respect to the proposed criteria, assessments of the quality of evidence used to inform judgements about the effects of interventions, and summaries of research evidence of the effects of health system interventions.

Methods

We conducted an international online survey to determine perceptions of the importance and use of the criteria within the DECIDE framework and evidence summaries. We aimed to survey a diverse (rather than representative) group of policymakers from different countries, with a wide range of experience with different types of health policy and management decisions and with different perspectives. Our sampling frame included the nine DECIDE partner countries (Canada, England and Wales, Finland, Germany, Italy, the Netherlands, Norway, Scotland, and Spain), the World Health Organization (WHO), and six SURE (Supporting the Use of Research Evidence for policy in African health systems) partner countries, namely Cameroon, Ethiopia, Mozambique, Zambia, Uganda, and South Africa [www.

who.int/evidence/sure/en/]. SURE is a related collaborative research project with the objective of developing and evaluating strategies for improving access to and use of research evidence in health policy development in Africa. These countries were included as we had sufficient networks linked to the DECIDE and SURE projects to provide the necessary referrals to potential participants. Inclusion criteria for survey participants included people responsible for health system decisions and stakeholders with an interest in and experience with health system decisions, such as delivery, governance and financial arrangements, and strategies for implementing health system changes. This could include public officials, managers, health workers, and representatives of special interest groups, international organizations, non-governmental organisations, donor countries, or the general public. Our partners in the DECIDE and SURE projects from each country and other personal contacts helped us to identify 10 to 20 people who were not involved with the DECIDE collaboration in each of the 15 countries and within WHO, with the aim of obtaining 10 completed questionnaires from each. We prepared an online survey (in English only) using the LimeService online platform [https://www.limeservice.com/] (Additional file 1 Table S1). The survey was revised after pilot testing in a small group of policymakers who provided feedback on the content, length, clarity, and ease of use. Informed consent was obtained from survey participants prior to commencing the survey and results were de-identified when exported for analysis to protect confidentiality.

To determine current use and perceived importance of the DECIDE criteria in health system decision-making, we asked participants to provide an example of a health system decision with which they had been involved or were familiar. Participants were asked to rate whether each of the ten criteria in the DECIDE framework (seriousness of the problem, number of people affected, quality of available evidence, benefits (desirable effects), adverse (undesirable) effects, resource use (cost), value for money, impacts on equity, implementability (feasibility), and acceptability) had been considered as part of this healthcare decision on a 3-point scale (yes, no, unsure). If a given criterion had not been considered, they were asked whether they considered it relevant, and also to state any other criteria they considered relevant. The opportunity to make optional comments was provided. Participants were also asked to rate the importance (important, probably important, not sure, probably not important, not important) of each of the ten criteria, with optional comments if desired.

To determine perceptions on the use of evidence in health system decisions, we asked participants about their use of 1) evidence from other countries; 2) systematic reviews to inform their health system decision; and 3) evidence grading systems. Specifically, we asked whether they believed that evidence grading systems should be the same or different for health system decisions, compared to clinical decisions. We also assessed the perceptions of participants as to the contents of summaries of evidence. These summaries, based on the best available evidence, should be concise yet include all of the key information needed to inform a health system decision. We asked participants to rate the importance of including certain types of information (components) in these summaries, namely: effect sizes, confidence intervals, numbers of studies, and the quality of the evidence. Participants also provided information on their research, healthcare, and decision-making experience.

Participants were contacted by email and asked to complete the online survey. Initial contacts were made by our partners in each country or directly by us. Non-responders received reminders via email at two and four weeks after the initial invitation. We summarised the results using frequencies and percentages, and collated provided comments. Our primary analysis focused on implications for our evidence to health system decisions framework and evidence summaries. We explored potential differences in responses across participants from different countries (DECIDE *versus* SURE partner countries) and across groups with different types of experience (with *versus* without research training or experience).

Results

We received 112 responses (70% response rate) to the online survey. Of these, 84 responses (75%) were complete and 28 (25%) were partially complete. We received 23 responses (46% response rate) from the six SURE countries, 84 responses (93% response rate) from the nine DECIDE partner countries, and five responses (50% response rate) from WHO. Table 1 describes the characteristics of survey participants who provided background information (n = 84). Most (84.5%) of the respondents had healthcare professional training or had worked as a healthcare professional. The majority of these were physicians (87.1%), and most (69.0%) had over ten years clinical experience. Most (78.6%) also had some form of research training or experience; 54.5% of these had over ten years of research experience. Respondents had worked in a variety of organizations, most frequently in national governments (63.1%) and public organizations (48.8%). The most common forms of health system decisions made or supported by respondents were decisions regarding the selection of healthcare policies, reforms or programmes (77.4%), and decisions on their implementation (78.6%).

Table 1 Characteristics of survey participants

	Survey participants, n (%)
Total (completed entire survey)	84 (75.0)
No healthcare professional training or experience working as a healthcare professional	13 (15.5)
Healthcare professional training or experience working as healthcare professional	71 (84.5)
Physician	61 (85.9)
Nurse	1 (1.4)
Other	8 (11.3)
Missing	1 (1.4)
Years of experience as healthcare professional	
1 – 10 years	22 (31.0)
11 – 20 years	18 (25.4)
21 – 30 years	14 (19.7)
31 – 40 years	14 (19.7)
40+ years	3 (4.2)
Research training or experience working as a researcher	66 (78.6)
MSc or equivalent	24 (36.4)
PhD or equivalent	40 (60.6)
Other	2 (3.0)
Years of experience as a researcher	
1 – 10 years	30 (45.5)
11 – 20 years	20 (30.3)
21 – 30 years	12 (18.2)
31 – 40 years	3 (4.5)
40+ years	1 (1.5)
Levels of current and previous work*	
International governmental organization	15 (17.9)
National government	53 (63.1)
Regional government	26 (31.0)
Local government	18 (21.4)
Public organization	41 (48.8)
Private organization	16 (19.1)
Other	19 (22.6)
Types of health system decisions supported	
Experience representing the views of stakeholders in policy or management processes	46 (54.8)
Decisions about healthcare policies, reforms or programmes	65 (77.4)
Decisions about implementation of healthcare policies, reforms or programmes	66 (78.6)
Management decisions about health system arrangements	44 (52.4)
Decisions about quality improvement, patient safety or implementation of clinical guidelines	56 (66.7)
Other	5 (6.0)

*These categories were not mutually exclusive and hence the sum is greater than 100%.

We asked participants to describe a recent healthcare decision with which they had been involved or had the opportunity to follow closely. One hundred and two respondents (91.1%) had such an example and 10 (8.9%) respondents imagined a current or recent decision. Figure 1 shows their responses regarding whether our criteria were considered for their health system decisions. Over 75% of respondents stated they had considered these criteria in their decision, except for value for money (67%) and impacts on equity (70%). Comments were infrequent but almost universally in support of the need to consider these criteria and the lack of evidence on certain criteria, such as equity and

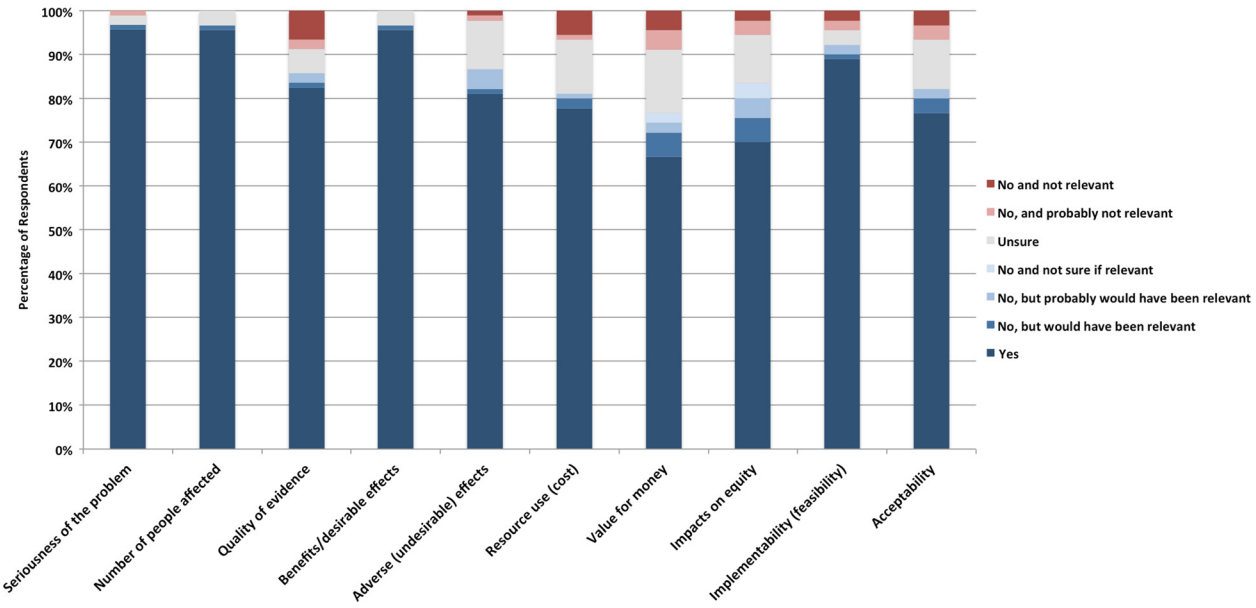


Figure 1 Survey responses regarding criteria considered for individual health system decisions.

implementation, for health system decisions. Two respondents indicated that consideration of these criteria is mandatory for decisions taken in their organization. Most (86.4%) of the respondents had considered evidence from other countries for their decision, and 94.3% of respondents agreed or somewhat agreed that evidence from other countries should be used to inform health system decisions. Most (90.9%) of the respondents stated they knew what a systematic review was. However, only 60.2% of respondents had used evidence from a systematic review to inform their decision. In response to our question on whether systematic consideration of the available research evidence would help to improve health system decision-making, 98.9% stated it would or it probably would.

Respondent ratings on the importance of the ten criteria are summarised in Figure 2. Every criterion was rated as important or probably important by over 90% of respondents. Nearly all respondents (99.2%) agreed that explicit consideration of the ten criteria would help or probably help to improve health system decision-making. While comments supported the importance of the criteria, several respondents identified further criteria. These included sustainability of implemented changes, post-implementation monitoring systems, human resource

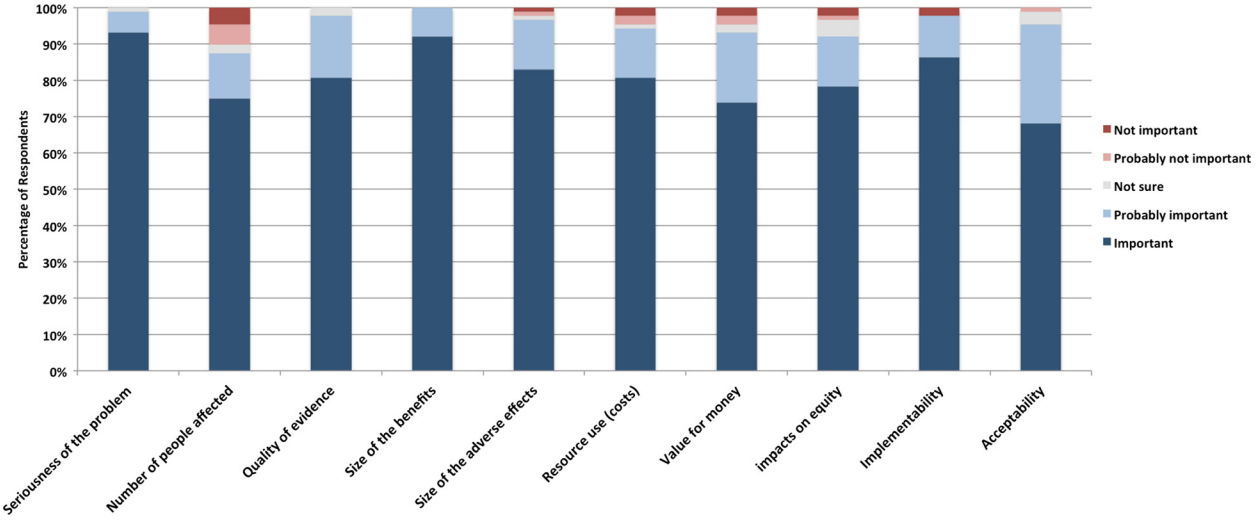


Figure 2 Survey responses rating the importance of criteria in health system decisions.

implications, and "environmental" considerations (assessing what is occurring in nearby areas or similar jurisdictions).

Most respondents agreed that a system of grading evidence would either definitely (59.7%) or probably (33.3%) improve health system decision-making processes (Table 2). When asked whether such grading systems should be the same for clinical and health system decisions, 38.4% said it should or probably should be the same, 45.3% indicated that different grading systems should or probably should be used for clinical and health system decisions, and 16.3% were neutral. This disagreement was also reflected in comments. Some respondents stated that using the same grading system improves consistency, transparency, and reproducibility, while others stated that they are fundamentally different types of decisions and evidence, requiring different grading systems. Two respondents (both from DECIDE partner countries) indicated an overall dislike of evidence grading systems. Respondent ratings on the importance of components of a summary of evidence are described in Figure 3. There was general agreement that all of the summary components were important or probably important, ranging from 79% (confidence intervals for effect estimates) to 96% (description of the quality of the evidence). All comments were in broad support, including several highlighting the need for clarity, simplicity, and brevity. Others suggested the use of disability-adjusted life years (DALYs) and number needed to treat (NNT) as important measures of effect.

Overall, there was relatively little variation in the responses. Nonetheless, we explored potential differences in responses across participants from different countries and across groups with different types of research experience (Additional file 1 Table S1). We did not find any apparent differences in responses based on the respondents' country (DECIDE *versus* SURE partner countries) or experience (with compared to without research training or experience).

Discussion

We received 112 responses to an online survey from a diverse, international group of policymakers, managers, their support staff and other stakeholders to better understand their perceptions of the ten criteria in the DECIDE framework for going from evidence to health system decisions. We also obtained their perceptions of the components of summaries of evidence. Our respondents had a high level of professional healthcare and research experience and training and had worked in a wide range of levels and organizations. This may be due to the way in which they were identified, or a greater interest among those policymakers with clinical or research experience to participate in the survey. Our sample of policymakers was not intended to be representative of included countries, but was rather to obtain a diverse range of views from respondents.

Respondents had generally considered all ten criteria in the DECIDE framework in their own healthcare decisions and consistently rated these criteria as important or probably important to decision-making. In earlier work, Guindo and colleagues identified a range of healthcare decision criteria and criteria-based decision-making tools used in empirical studies of health system decisions – the most frequently cited criteria were equity and fairness, efficacy/effectiveness, stakeholder interests and pressures, cost effectiveness, strength of evidence, and safety [16]. In our survey, comments by respondents indicated that not only were similar criteria considered

Table 2 Respondents' views on the use of grading systems to rate the quality of evidence of health decisions

	Survey participants, n (%)
A system of grading the quality of evidence can help improve health system decision-making processes.	
Yes	52 (59.7)
Probably	29 (33.3)
Not sure	5 (5.8)
Probably not	1 (1.2)
No	0 (0.0)
Do you think that a system for rating the quality of evidence should be consistent for different types of decisions or that there should be different systems for different types of decisions (e.g., for clinical decisions and for health system decisions)?	
The SAME SYSTEM DEFINITELY should be used for rating the quality of evidence for clinical and health system decisions	9 (10.5)
The SAME SYSTEM PROBABLY should be used for rating the quality of evidence for clinical and health system decisions	24 (27.9)
Neutral	14 (16.3)
A DIFFERENT SYSTEM PROBABLY should be used for rating the quality of evidence for health system decisions than for clinical decisions	27 (31.4)
A DIFFERENT SYSTEM DEFINITELY should be used for rating the quality of evidence for health system decisions than for clinical decisions	12 (13.9)

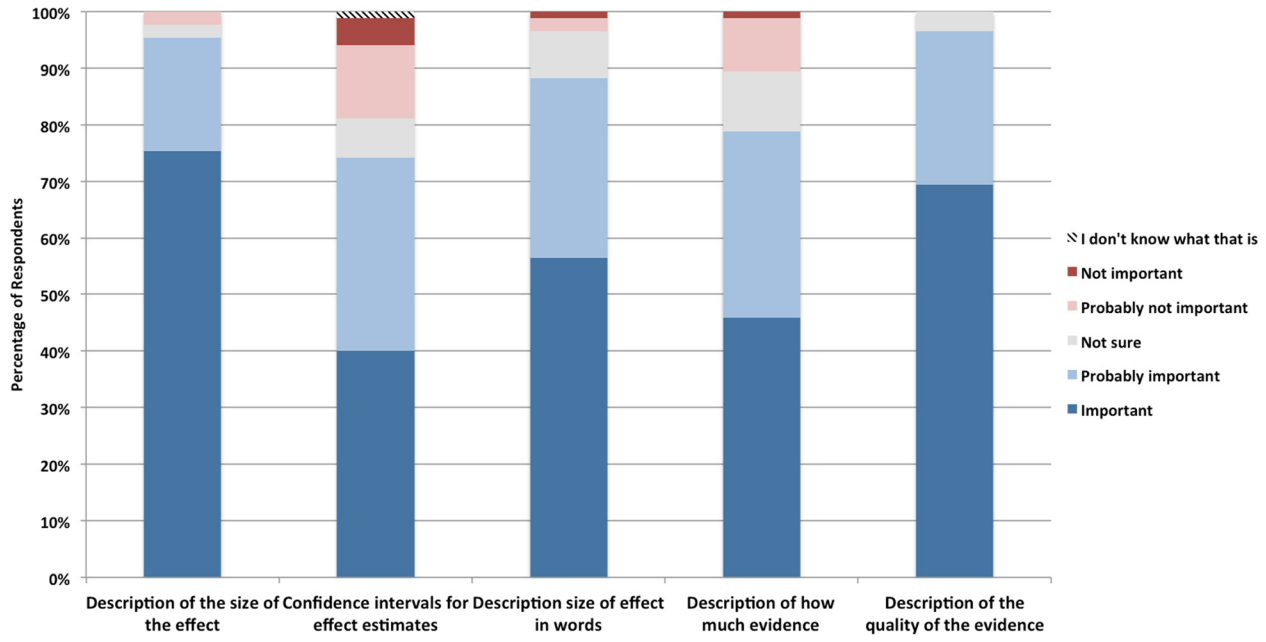


Figure 3 Survey responses rating the importance of components of a summary of evidence.

important to decision-making processes, but that in some organizations, their consideration is mandatory. However, only 60.2% reported using systematic reviews to inform decision making. This may reflect both a lack of systematic reviews addressing relevant questions, as well as other reasons for not using research evidence to inform health policy decisions, as reported elsewhere [17,18].

Most respondents agreed that a system of grading evidence would improve health system decisions; however, there was considerable disagreement as to whether these grading systems should be uniform across clinical and health system decisions. This is consistent with the ongoing debate in the literature on the application of the GRADE approach to a range of decisions [10,12]. While a common approach to grading evidence may reduce confusion, minimise conflicts of interest and enhance intellectual rigor, others have argued that this approach can be overly complex and can favour false negative conclusions. Additionally, interventions that are not amenable to randomised trials could be disadvantaged in terms of prioritisation, funding and implementation [12]. Criticisms made by our respondents of a uniform grading approach concur with the literature that there is a lack of available high-quality evidence on health system interventions. Two respondents also indicated an overall dislike of grading systems, citing a tendency to oversimplify complex issues or a lack of institutional capacity to provide training or support grading activities. However, the application of some form of systematic consideration of evidence to support health system decisions had broad support in our survey. Successful implementation of the GRADE approach for health system decision-making will need to address concerns about its applicability to health system evidence.

Effectively communicating complex information through summaries has been proven in the use of the GRADE summary of findings table [13,19] and SUPPORT summaries [2]. SUPPORT summaries summarise the best available evidence of the effects of health system interventions for low and middle-income countries. Such summaries may be particularly useful for policymakers without a strong health or research background. Although our respondents had a higher-than-expected level of healthcare and research experience, their responses clearly indicated that all proposed components of the summaries of evidence have practical applications in health system decision processes. Comments strongly favoured clear, concise summaries in simple language, suggesting that researchers should consider ease of interpretation by policymakers with limited scientific literacy when preparing summaries.

The strengths of this survey were a good response rate from a diverse range of countries, backgrounds, levels of decision-making, and organisations. The fact that 25% of the surveys were partially completed may have been due to survey length. The comparatively lower response rate from SURE partner countries means results are likely biased towards higher-resource settings, limiting their applicability to resource-constrained settings. One significant limitation of the survey was the relatively high level of healthcare and research experience amongst respondents; only four respondents had no healthcare or

research experience. This is probably not representative of health policymakers in general and potentially limits the generalizability of our findings. Nonetheless, the fact that the criteria included in the DECIDE framework were regarded as important for decision-making by these respondents provides further support for their inclusion in a framework for going from evidence to health system decisions.

This survey confirmed the relevance of the criteria that we had identified and incorporated in the DECIDE framework for health system decisions and suggests that the framework is likely to be helpful for informing health system decisions. Further development and evaluation of the framework will be based on practical applications of the framework to health system and population health decisions and user testing [20]. Facets of the framework that will be addressed by user testing were adapted from the work by Morville et al. [21] and Rosenbaum et al. [22], and include 1) findability: can users locate what they are looking for?; 2) usefulness: does the framework have practical value for the user?; 3) usability: how easy and satisfying is the framework to use?; 4) understandability: do users understand the framework and the content correctly?; 5) credibility: is this framework/content trustworthy?; 6) desirability: is the framework something the user wants/responds positively to?; and 7) identification: does the framework feel like it was designed for "someone like me (the user)"? Further work will address the advantages and disadvantages of using the same *versus* different systems for grading evidence, further clarification of the included criteria, the need for additional criteria, and the perceptions of policymakers and stakeholders who do not have a research or health professional background.

Conclusions

Health system decision-making requires careful consideration of a multitude of variables, such as the magnitude of the problem, the size of benefits and adverse effects, feasibility and acceptability, as well as resource and equity implications. Surveyed individuals supported the use of systems to grade the quality of evidence for health system decisions, but there was disagreement as to whether uniform or different grading systems should be used for health system and clinical decisions. Communication of evidence to policymakers and stakeholders involved in health system decisions should employ succinct summaries of measures of effect and the quality of evidence in clear and simple language.

Additional file

Additional file 1: Table S1. Participant responses on quality of evidence and components of evidence summaries, stratified by research experience and country of origin. Differences in responses by respondent research experience and country of origin were tested using χ^2 tests; none were significant at the P <0.05 level.

Abbreviations

DALY: Disability-adjusted life years; DECIDE: Developing and evaluating communication strategies to support informed decisions and practice based on evidence; GRADE: Grading of recommendations assessment development and evaluation; HRP: Special programme of research development and research training in human reproduction; NNT: Number needed to treat; SUPPORT: Supporting policy relevant reviews and trials; SURE: Supporting the use of research evidence for policy in African health systems; UNDP: United nations development programme; UNFPA: United nations population fund; WHO: World health organization.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

All the named authors contributed to the survey design and content. JPV and ADO administered the survey and JPV conducted the analysis. JPV and ADO prepared the initial draft of the manuscript and all authors contributed to and approved the final version of the manuscript. This manuscript represents the views of the named authors alone. All authors read and approved the final manuscript.

Acknowledgements

The research was funded in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 258583 (DECIDE project). JPV's position at the World Health Organization is funded by an Australian Postgraduate Award and the A&A Saw Scholarship.

Author details

[1]School of Population Health, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia. [2]UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP) / Department of Reproductive Health and Research, World Health Organization, Avenue Appia 20, Geneva 1211, Switzerland. [3]Global Health Unit, Norwegian Knowledge Centre for the Health Services, P O Box 7004, St. Olavsplass, 0130, Oslo, Norway. [4]Health Systems Research Unit, Medical Research Council of South Africa, PO Box 19070, 7505, Tygerberg, South Africa.

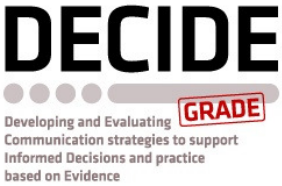
Received: 1 November 2012 Accepted: 20 May 2013
Published: 24 May 2013

References

1. Glenton C, Santesso N, Rosenbaum SE, Nilsen ES, Rader T, Ciapponi A, Dilkes HH: **Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study.** *Med Decis Making* 2010, **30:**566–577.
2. Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, Young T, Althabe F, Ciapponi A, Marti SG, Meng Q, Wang J, La Hoz Bradford AM, Kiwanuka SN, Rutebemberwa E, Pariyo GW, Flottorp S, OA D: **Evidence summaries tailored to health policy-makers in low-and middle-income countries.** *WHO Bull* 2011, **89:**54–61.
3. Lavis JN, Permanand G, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP) 13: Preparing and using policy briefs to support evidence-informed policymaking.** *Health Res Policy Syst* 2009, **7**(Suppl 1):S13.
4. The SURE Collaboration: *SURE Guides for Preparing and Using Evidence-Based Policy Briefs.* 2nd edition. Oslo, Norway: SURE Collaboration; 2011.
5. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L: **Systems to Rate the Strength of Scientific Evidence: Summary.** In *AHRQ Evidence Report Summaries.* Rockville (MD): Agency for Healthcare Research and Quality (US); 2002.
6. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, Group GRADE Working: **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.** *BMJ* 2008, **336:**924–926.

7. GRADE working group: *Organizations that have endorsed or that are using GRADE*. Available from: [http://www.gradeworkinggroup.org/society/index.htm].
8. 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:**35–43.
9. European Centre for Disease Prevention and Control: *Evidence-based methodologies for public health – How to assess the best available evidence when time is limited and there is lack of sound evidence*. Stockholm: ECDC; 2011 [http://ecdc.europa.eu/en/publications/Publications/1109_TER_evidence_based_methods_for_public_health.pdf].
10. Lewin S, Bosch-Capblanch X, Oliver S, Akl EA, Vist GE, Lavis JN, Ghersi DD, Røttingen J-A, Steinmann P, Gulmezoglu M, Tugwell P, El-Jardali F, Haines A: **Guidance for evidence-informed policies about health systems: assessing how much confidence to place in the research evidence.** *PLoS Med* 2012, **9:**e1001187–e1001187.
11. Rehfuess EA, Akl EA: **Current experience with applying the GRADE approach to public health interventions: an empirical study.** *BMC Publ Health* 2013, **13:**9.
12. Schünemann HJ, Fretheim A, Oxman AD: **Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations.** *Health Res Policy Syst* 2005, **4:**21.
13. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbegovic B, Alonso-Coello P, Post PN, Busse JW, Glasziou P, Christensen R, Schünemann HJ: **GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes.** *J Clin Epidemiol* 2013, **66**(2):158–172.
14. Rosenbaum SE, Glenton C, Nylund HK DOA: **User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews.** *J Clin Epidemiol* 2010, **607:**619.
15. Rosenbaum SE, Glenton C, Oxman AD: **Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information.** *J Clin Epidemiol* 2010, **63:**620–626.
16. Guindo LA, Wagner M, Baltussen R, Rindress D, Van Til J, Kind P, Goetghebeur MM: **From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking.** *Cost Eff Resour Alloc* 2012, **10:**9.
17. Innvaer SS, Vist GG, Trommald MM, Oxman AA: **Health policy-makers' perceptions of their use of evidence: a systematic review.** *J Health Serv Res Policy* 2002, **7:**239–244.
18. Lavis J, Davies H, Oxman A, Denis J-L, Golden-Biddle K, Ferlie E: **Towards systematic reviews that inform health care management and policy-making.** *J Health Serv Res Policy* 2005, **10**(Suppl 1):35–48.
19. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, Group GRADE Working: **What is "quality of evidence" and why is it important to clinicians?** *BMJ* 2008, **336:**995–998.
20. *Improving the user experience of evidence: A design approach to evidence-informed health care*. Oslo School of Architecture and Design; 2010 [http://nokc.academia.edu/SarahRosenbaum/Papers/369198/Improving_the_user_experience_of_evidence._A_design_approach_to_evidence-informed_heath_care].
21. Morville P, Semantic Studios: *User Experience Design*; 2004 [http://semanticstudios.com/publications/semantics/000029.php].
22. Rosenbaum SE, Glenton C, Cracknell JJ: **User experiences of evidence-based online resources for health professionals: user testing of The Cochrane Library.** *BMC Med Inform Decis Mak* 2007, **8:**34.

doi:10.1186/1478-4505-11-19
Cite this article as: Vogel *et al.*: Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey. *Health Research Policy and Systems* 2013 **11**:19.



Appendix 2

Advisory group comments regarding the evidence to decision framework and our responses

Appendix 2. Summary of advisory group feedback on the evidence to decision framework

15/32 feedback forms returned plus one set of comments without a form, ie feedback from 16 people in total (50%)

Initials indicate who the comments were made by

				Comments	Response
Comprehensiveness **1.** Are there important considerations that are missing from the framework?	Yes 9	Uncertain 1	No 5	In practice, the government policy development that we support tends to result in complex, multidimensional strategies and policies that involve multiple interventions. For example, the recently announced Ontario comprehensive mental health and addictions strategy (MHA) has many different elements including introducing new types of supports and care (e.g., education, early detection), through new providers in new settings. The MHA strategy is an extreme example, but most government strategies/policies involve more than one intervention. Our experience is that this also is the case when stakeholders advocate for policy change - they rarely ask for one thing. That being the case, it would be very hard for decision makers or policy development staff to fill out the evidence column of the framework, since each strategy/policy is complex and draws on different evidence: systematic reviews, RCTs, qualitative research, studies of administrative data etc. (AP)	Separate components of a strategy can be summarized and assessed separately (e.g. for the mental health strategy) or a bundled package of components can be assessed, if they are to some extent interdependent. Ask for an example of how they would currently present complex evidence for a complex option. Alternatives are to summarise the key evidence in a structured (or unstructured) way or to leave it up to decision makers to wade through a pile of papers without the help of a structured summary. Ask if he would argue against a structured summary or would advocate a different structure (and, if so, what).
				We see the framework as being a very useful tool if used more than once as part of an iterative policy development process. It prompts people to think of all the things they are trying to maximize and minimize with their strategy/policy. If there is a desire to have	Agree it can be used in that way, but does not necessarily need to be used in that way. Agree that there is need for text that describes the purpose of the framework and ways in which it can be used.

	Comments	Response
	it used this way, we'd suggest adding some text to that effect at the front. (AP)	
	Might want to include an element related to timing of implementing a decision? (ML)	Add timing to the information at the top? In text describing the framework include how time needs to be considered in judgments for several criteria (rather than as an independent criterion). See above. Outline different ways in which time enters into consideration of different factors: when benefits and costs occur in relationship to acceptability, durability/sustainability of the intervention in relation to implementation, etc.
	On page 2, It appears to me that an additional factor determining the importance of a health policy or program is the projected time frame. For many managers, policies that bring short-term benefits are preferable to policies that bring long-term benefits. This factor may be particularly important for managers/decision makers working in a setting of uncertain funding, in which implementing short-term policies may be preferable because they draw upon current funds, while long-term policies may rely upon uncertain future funding. This factor may also be important to managers who are elected or seeking reappointment within a certain time. (A)	
	-Trends - Social Relevance (XC)	There maybe should be a prompt to consider trends in relationship to the size of the problem (and baseline in the SoF table). Not clear what is meant by social relevance. Check with XC. Correct. We are not suggesting grading this evidence at this point in time. Prompts can provide guidance

	Comments	Response
	I think most of the relevant factors are in the proposed framework. Seriousness and number of people affected deals with questions of burden of disease. Epidemiological studies are needed to get such information and they will always be observational, and will be classified as low or best as moderate evidence in a GRADE grading. You have not applied such grading for those criteria, and that is probably wise, but the need for good evidence for these questions is as important as for the intervention questions.	for taking uncertainty into consideration when presenting the evidence and making a judgement about this.
	Urgency of a health problem is another relevant criterion in public health, especially in situations of infectious disease outbreaks.	Add prompt to consider this in relationship to seriousness (+ with adding timing at the top)
	Other possible criteria to consider: • Disagreement among stakeholders about important interventions • Ethical considerations in a wider concept than only equity considerations • Considerations related to local context, sex, jurisdiction (FF)	Add prompt to consider this prior to preparing the framework and possibly in the background at the top. This is part of acceptability. Not clear what he has in mind. Check with FF. Local context is relevant to several criteria. Sex may be an equity issue, may be an applicability (quality of evidence) issue. Jurisdiction may be an implementation issue (authority and accountability). They are stakeholders.
	It is a comprehensive framework. I do wonder if acceptability should include not only the acceptability of the stakeholders but also of the decision makers themselves. If civil servants are pulling the information together for Ministers for example, there will potentially be options that Ministers will not agree with but when weighed up against the other costs/benefits it may work out best option for that Ministry.(LH)	Clarify that this is part of certainty of the expected effects.
	Under types of decision it might be appropriate to say something about the applicability of the data (LV)	See above.

	Comments	Response
	You could consider including a consideration of 'Sustainability'. This criterion refers to the durability of the intervention considering factors such as: • The level of ongoing funding support required • The community empowerment and capacity building required and level of policy support likely to be achieved • The likelihood of required changes in behaviours, practices and attitudes being achieved on an ongoing basis. *Note: the last 2 dot points are more relevant to health promotion/public health interventions.* References: Haby and Bowen 2010, Swinburn et al 2005 (MH)	See above + suggest a two stage process.
	Policy-makers often have to make decisions about the amount they should invest, how extensively a given intervention should be implemented, or how fast, rather than yes/no decisions about whether or not to implement. Even highly cost-effective interventions may increase costs to the health service, so have to be weighed up in terms of opportunity costs. I don't think the framework will help with this kind of decision, unless it allows them to model and compare the consequences of a range of options . (PC)	They would be considered as another stakeholder.
	For countries dependent on significant external funding, considerations would also need to cover the decision making influence of donor partners/agencies (S)	Agree that we need to add prompts/guides for each criterion.
	No, provided that criteria actually can contain many sub-criteira.(XBC)	

				Comments	Response
Relevance **2.** Are considerations included in the framework that should not have been?	No 13	Uncertain 1	Yes 1	Three of the criterion might be removed or changed. It is hard to imagine that someone would indicate that a problem is not serious if they have got to the point of considering new policy for it (criterion 1). The question about the number of people affected (criterion 2) would also be taken into account well before policy development starts, and in some cases there may be a decision to develop policy even though the numbers are small. Re: feasibility to implement (criterion 9), it seems unlikely that anybody would be developing policy that would get the answer "no" (See comment under weaknesses). (AP)	Keep. Problems are not all equally important. Feasibility may be an important criterion that drives a decision. There may be options that although there are problems with feasibility, it is important enough that the decision is to do something, taking account of those considerations.
				'Seriousness of the problem' and 'Number of people affected' should be combined into one criterion to reflect the size of the problem. Total DALYs, as measured in Burden of Disease studies would be better here because it combines N, severity and duration. Note: In the Assessing Cost-Effectiveness (ACE) studies, decision-makers found severity and size of the problem difficult to apply as a criterion (to distinguish between interventions) because it applies to the health problem rather than the interventions to address it. Reference: Carter et al 2008 (MH)	Keep for now. DALYs often may not be available or useful.
Applicability **3.** Is the framework applicable to different types of health system decisions?	Yes 10	Uncertain 5	No	See comment for Comprehensiveness, it would be difficult to fill out the evidence sections for complex, multi-dimensional strategies/policies. (AP)	See above.
				Not always., usually there is not enough scientific evidence related to the benefits of organizatived changes. (XC)	Still can be used (e.g. some of Optimize examples)
				I am not sure if this system is fitting well to make rapid risk assessments in public health. There is	Ask for examples. That may be true. We have not suggested it is useful for such decisions, although we

					Comments	Response
					something about preparedness and incident verification that is relevant to such processes that is not in here. Maybe better to define and narrow the scope of this framework? (FF)	are not sure how different a framework for rapid responses would be.
					As the questions are about the factors that should be taken into consideration when making a decision, it allows for transferability to different types of decisions. (LH)	Yes
					It would be helpful to add in more examples (LV) It does depend though on how the framework is implemented and supported. (MH)	Good suggestion.
					Yes, but may need to consider both formal and informal decision making processes (S)	Consider including this in explanation of how the framework can be used. See text in SUPPORT Tools.
4.	Is the framework applicable to different types of decision-making processes?	Yes 6	Uncertain 8	No 1	Often information on the size of effects of higher level strategies/policies (e.g. deciding whether to regulate a sector) is unavailable. This tool may be most applicable to operational/implementation policy. (AP)	See considerations apply and decisions need to be made regardless of what evidence is available.
					You might want to add more information as to the purpose of the tool and include definitions for abbreviated text. (ML)	Agree. See above.
					I am not sure if this system is fitting well to make rapid risk assessments in public health. There is something about preparedness and incident verification that is relevant to such processes that is not in here. Maybe better to define and narrow the scope of this framework? (FF)	
					I think it probably is but when it comes to weighing up several options when it has been decided that some form of action must be taken, if all of the options	Yes. This can be addressed by a higher level/matrix summarizing key information across multiple options.

			Comments	Response
			require to be postponed or piloted then the decision may not have been made any easier. Is there a way to then to weight the considerations and put them in a matrix for comparison for these types of decisions? (LH)	
			It does depend though on how the framework is implemented and supported. (MH)	Add something to explanation about this?
			Some, but not all (PC)	Ask which ones?
			Yes, but again may want to bring in some way to identify lack of participation, compliance, neglect, and laissez faire that leads to decisions not implemented. Barriers and incentives (S)	Yes, as implementation considerations.
			I would think that management decisions are of different nature (e.g more to do with the reallocation of resources than with the selection of strategies) (XBC)	Not clear. Ask XC if this means he thinks the framework would not work for management decisions.
Simplicity **5.** Is the framework more complicated than necessary?	No 12	Uncertain 3 Yes	It represents a good set of prompts. (AP)	Agree
			It is quite complicated, but policy making and public health decision making is complex and it is an aim to make the different implicit elements of the process and judgements explicit, as this tool does. Only by displaying the elements and the uncertainties connected to each element can awareness be raised and decisions better qualified. (FF)	Agree. Flag consideration of uncertainties with all relevant criteria.
			Some items might be combined, making fewer items to check. However, I like the way that the components have been broken out, and in general believe that it is better to err on the side of clarity and simplification rather than risk that something will be overlooked when combined into a composite	We agree it is better not to combine criteria. We are not sure what would be gained by combining the first two criteria.

				Comments	Response
				element. (e.g. the seriousness of the problem is obviously a combination of the health effect and the number of people affected but having them broken out makes sure that both are explicitly considered). Perhaps making the first question something like "Does the problem create serious health effects?" make the progression seem clearer and necessary. (MG)	
				This will depend on the level of support given in using the framework. (MH)	Agree.
				If anything, it's not complicated enough! (PC)	It is not clear how making it more complicated would help.
				Although it requires quite a systematic thinking (XBC)	Agree
Logic **6.** Is the framework organised in a logical way that is easy to understand?	Yes 13	Uncertain 2	No	Very straightforward and has a natural progression throughout the framework. (LH)	Thanks.
				I would reverse the order of the final two criteria. Having something that is acceptable to key stakeholders is key component to being able to implement. Much like # 5 above I like calling these out separately because the acceptability issue is something that could easily be overlooked by naive researchers or bureaucrats but if it comes first it makes more sense to me. (MG)	We agree and have switched this.
				Difficult to say in generic terms just as 'feasibility'; some may think that costs are part of feasibility (XBC)	We agree this may be a consideration and is included in our (SURE) checklist. (Compare SURE to TICD checklists. Add interactions.)
Clarity **7.** Are the criteria labelled and explained in a way that is easy to understand?	Yes 11 1 no answer	Uncertain 3	No	Criterion 6, "are the resources required relatively small" is open to interpretation. It may be helpful to add an endnote that provides guidance on what small and large mean in terms of dollar figures (perhaps as % of total budget), #of new providers required (perhaps as % of total workforce), see also	Agree we need guidance for all of the criteria. And some jurisdictions may want to develop their own guidance.

					Comments	Response
					comments for criteria 5 and 9 below under weaknesses. (AP)	
					The quality of evidence rating is using the GRADE terminology which is difficult to communicate when the best obtainable evidence is coming from studies called low quality evidence. For questions where observational studies, modelling, lab research or outbreak investigations are the best and only attainable evidence it is a problem to label it low quality. (FF)	Not clear how big a problem this is. Our survey may help to clarify this.
					But see comments below (LV)	
					However, page 2 is needed to make sure that criteria such as "Are the desirable effects large" or "Are the undesirable effects small" etc. Are clearly understood (MG)	Agree.
					I assume you're referring here to the questions in the template (and 2 examples). On their own they are quite subjective, but with appropriate support and explanation they will become more objective. For example, 'Is the problem serious?' How is serious defined? Are there benchmarks that the user can use to make it easier to answer objectively? (MH)	See above.
					Needs thinking; but probably not obvious to find better ways (XBC)	OK.
Usability **8.**	Would it be easy for people responsible for health systems decisions to use the framework?	Yes 7	Uncertain 7	No 1	We think it would be difficult to fit the evidence for multidimensional strategies/policies into the framework. We would see the value of using it (more than once) to prompt refinements and improvements during an iterative policy development process. (AP)	See above.

	Comments	Response
	Good to have guiding information on what the can help decision makers accomplish (ML)	Agree.
	Depens, some groups of " staldeholders" need more explanations on some concepts. (XC)	Agree.
	The framework is logic and understandable, but needs good interpreters (FF)	Agree.
	Assuming, as mentioned they can get access to the evidence and easy access to thing like SUPPORT tools etc.(LH)	Agree.
	While there will be elements of the information included in a well completed template that might be hard for some to understand, the concept, layout, and organization is easy to understand and to use. I think some of the challenges will be in clearly communicating research especially on things such as problem and effect sizes. (MG)	Agree (and thanks from Sarah and Angela)
	Only if well supported with the research evidence, benchmarks etc to enable them to implement the framework. If not well supported, the decisions will end up being very subjective. Also, even if the framework is well used, the final assessments of 'Balance of consequences' and 'Decision' will still be difficult as it is difficult to balance all of the criteria. For example, even if the evidence is strong, the intervention is cost-effective and the health impact large but it has a negative impact on health inequalities or the workforce is not available, how does a decision-maker decide what to do? What is the 'right' decision? (MH)	Agree.
	I expect that the kinds of policy-makers I deal with (in the Scottish Government Health Directorates) would	We agree and plan on doing this.I

				Comments	Response
				be able to use the framework, if it could be applied to the kind of decision-making problems they deal with. However, this and the next four questions could only be answered though a pilot or trial run of some kind (PC)	
				This question should be empirically tested (RG)	Agree.
				As in response to 1. above, people responsible for decisions also ought to include external decision making influence such as donors. Framework may benefit from bringing in perspectives from the "aid effectiveness" principles in this regard and also assess the match between principles and reality when it comes to who owns the decisions (S)	Donors can be included as stakeholders and may be relevant to consider in relation to implementation. Check with S re aid effectiveness principles and what she means.
				Depending on the background. It may not be easy, but still worthwhile (XBC)	Agree.
9.	Would it be easy for stakeholders to use the framework?			We think this tool would be useful for stakeholders, but they may not have access to information to answer all the questions (e.g., resources, cost, implementation considerations) (AP)	Agree.
		Yes 7	Uncertain 7	No 1	Only if well supported with the research evidence, benchmarks etc to enable them to implement the framework. If not well supported, the decisions will end up being very subjective. Also, even if the framework is well used, the final assessments of 'Balance of consequences' and 'Decision' will still be difficult as it is difficult to balance all of the criteria. For example, even if the evidence is strong, the intervention is cost-effective and the health impact large but it has a negative impact on health inequalities or the workforce is not available, how does a decision-maker decide what to do? What is the 'right' decision? (MH)	See above. If, for example, there is a concern about impacts on inequities and there are is no evidence, that can be flagged as something to monitor.

				Comments	Response
				I expect that the kinds of policy-makers I deal with (in the Scottish Government Health Directorates) would be able to use the framework, if it could be applied to the kind of decision-making problems they deal with. However, this and the next four questions could only be answered though a pilot or trial run of some kind (PC)	See above.
				This question should be empirically tested (RG)	See above.
				Yes, but since there is no obvious way to bring in and deal with conflicting interests and priorities, it has some limitations.I like the part on relevant factors, pros and cons of complex options, but need to see how it works out in reality (S)	Good point. Consider ways of addressing competing interests in the introduction / description of the framework. Competing interests of others are addressed under acceptability.
				Depending on the background. It may not be easy, but still worthwhile(XBC)	See above.
Suitability **10.** Is the framework suitable for informing and helping people to make health system decisions?				What about the training needs, technology requirements, monitoring & evaluation associated with a decision? Perhaps these could be included in the framework? (ML)	Good point. We will consider this.
	Yes 11	Uncertain 3	No 1	Yes, but it will be most helpful if well supported, e.g. by those able to interpret and analyse the research evidence. (MH)	Agree.
				I expect that the kinds of policy-makers I deal with (in the Scottish Government Health Directorates) would be able to use the framework, if it could be applied to the kind of decision-making problems they deal with. However, this and the next four questions could only be answered though a pilot or trial run of some kind (PC)	See above.

				Comments	Response
				This question should be empirically tested See also comments below at item 16 (RG)	See above.
				Suitable as one among several tools (S)	Ask S what other tools she has in mind.
Usefulness **11.** Is the framework likely to be useful to people responsible for health systems decisions?	Yes 9	Uncertain 5	No 1	I would think so in terms of its helpfulness but the issue is more likely to be in getting the right people to know about it to start using it. Getting buy in from the right people at the right level so that they expect decisions to be made using the framework. (LH)	Good suggestion.
				I think more case studies are needed. My instinct is that it will be useful (LV)	Agree.
				However it depends on what their agenda is. If they are in the US and they are looking to maximize profit at the expense of other considerations, the framework could hinder their scheme. On the other hand, if the data supported their instincts then it would be helpful. Most importantly, if their objective was to make well informed policy decisions then clearly this is helpful. (MG)	Thanks. Could be used to make explicit that profit maximization is the goal.
				I expect that the kinds of policy-makers I deal with (in the Scottish Government Health Directorates) would be able to use the framework, if it could be applied to the kind of decision-making problems they deal with. However, this and the next four questions could only be answered though a pilot or trial run of some kind (PC)	See above.
				This question should be empirically tested (RG)	See above.
				Depending on its ability to pick up on some of the comments above (S)	Ok

				Comments	Response
				Probably not by itself, but within an evidence-informed decision-making culture. (XBC)	Might also help to develop such a culture.
12. Is the framework likely to be useful to stakeholders?				Might want to pilot this with a stakeholder? (ML)	Agree.
				Same point as #11 above. It depends on what they are trying to accomplish. If they are an advocate for an enormously expensive intervention that is marginally helpful in a rare disease then it won't be terribly helpful. If they are trying to advocate for effectively allocated scarce resources then it would be. (MG)	Agree.
	Yes 7	Uncertain 7	No 1	I expect that the kinds of policy-makers I deal with (in the Scottish Government Health Directorates) would be able to use the framework, if it could be applied to the kind of decision-making problems they deal with. However, this and the next four questions could only be answered though a pilot or trial run of some kind (PC)	See above.
				This question should be empirically tested (RG)	See above.
				Depending on its ability to pick up on some of the comments above (S)	
				Probably not by itself, but within an evidence-informed decision-making culture. (XBC)	See above.
Overall assessment **13.** Overall, is the framework adequate for its intended purpose?				It is a helpful tool for people who are making health system decisions to consider all the important factors when making decisions. (LH)	Thanks.
	Yes 8	Partially 6	No 1	I think some guidance about how long it takes to complete the data for this would useful (LV)	Good suggestion. Refer to SUPPORT Tool. Maybe focus on resources/capabilities that are needed.

	Comments	Response
	The main modification suggested, as mentioned above, would be to combine the first two criterion into one. It will also need lots of support. But overall, it's a good framework. (MH)	See above.
Strengths **14.** What do you like about the framework?	This is a good series of prompts. We appreciate that there is not a score that implies a value for each component, leaving it to the analyst to determine which of the components should be weighted most heavily in the decision of whether to proceed. (AP) It highlights the important considerations. It provides a useful checklist for a decision-making process (AH) It's simple and not too long. The language is simple although it could be made even more plain and easy to follow.(EB) More detail on considerations when making a decision, goes beyond evidence analysis and adds more decision points, which is realistic. (ML) Introduction of systematic analysis. -They are explicits criterions. -The analysis is standardized for everyone (XC) The strength is the logic step by step procedure of assessing the different items of a decision making process and the attempt to make those elements explicit and as far as possible quantifiable. At the same time not all elements can always be quantified, and it is legitimate to make judgments based on values, ethics, experience and consensus as long as	Thank you.

	Comments	Response
	it is done openly and explicitly, and the reasons for doing it is stated. Expressing effect estimates in Risk differences is easier to capture and more logic, I think than the use of relative risk reduction or differences in Odds Ratios (FF) I like its simplicity. This means people won't be put off using the framework time and again.(LH) I think the report could be very useful to decision-makers to present data in a simple way. (LV) It is logical, inclusive of important factors, and clear (MG) Covers most relevant issues (MH) It is logically organised and puts the key evidence into the context of a decision problem. This is much more relevant and engaging than simply presenting summaries of effect sizes, ICERs, etc. (PC) The idea of disentangling the different possible components of the decision making process, thus making the process itself more transparent (RG) Takes a systemic and systematic view (S) It systematises and makes explicit the underlying thinking when decision on options. (XBC)	
Weaknesses **15.** What don't you like about the framework and what suggestions do you have for improving it?	The scale no to yes doesn't work well for all of the criteria. Perhaps consider changes for some, for example:	Will do.

	Comments	Response
	Criterion 9, 'is the option feasible to implement' might be improved if reworded as "risks and barriers to implementation" and the boxes could range from "many, successful implementation uncertain" to "few, easily overcome". (AP) For Criterion 5 "are the undesirable effects small" it would be helpful to ask users to comment on the degree to which undesirable effects can be mitigated, and if so, how. (AP) There may be many topics when many items are checked "uncertain" but that is due to a lack of evidence, not a weakness of the framework. (AH)	
	This tool only helps policy makers reach a decision when the evidence is overwhelming, in which case the decision is obvious anyway. If there is any doubt then the 'balance of consequences' concludes that a decision should be postponed, evaluated or a pilot study should be carried out. For this reason I have to conclude that the tool will not be of use to policy makers in its present form.(EB)	We disagree, but will explore and test this. Evidence does not make decisions obvious and the framework may help when there is little evidence.
	The framework also omits a crucial question: How reliable is the evidence? This is a different question from the one posed, 'is there high or moderate quality evidence (little or moderate uncertainty) about the size of the desirable and undesirable effects of the option'. The	Not clear what is meant by 'reliable evidence'. (May need to clarify how this is considered in the framework.)

	Comments	Response
	framework does not encourage the user to ask questions about the reliability of the evidence. Who conducted it? What was the sample size? Did they have any conflicts of interest? (EB)	
	It also ignores some of the complexities of translating evidence into action. In Example 1 (stroke care) the evidence clearly shows that patients have better outcomes when treated in stroke units. Several years ago Scotland announced its intention to do this. Yet many people are still not treated in specialist stroke units. Why? Because the initial investment isn't available. There is no analysis of whether the necessary infrastructure exists, or how long it might take to implement and what effect this will have on the final outcome. Currently there is a danger that the spreadsheet's conclusion merely repeats the evidence rather than provides a basis for policy.(EB)	This is a consideration under implementation.
	Could add something on the timing a decision requires ? Think this could be an important consideration when judging between one alternative and another? (ML)	Agree. See above. Agree, there often is missing evidence.

	Comments	Response
	- Most occasions the required information is not enough. Example: there are not information of prevalence or incidence from many diseases. Around the world there are just 8 studies of prevalence in cataracts. - Not enough evidence - Just Where there is evidence, actions are made - The method is not explicit when you compare two interventions. There are no scores. (XC)	See above about multiple options. We do not feel that is a basis for assigning scores.
	A comprehensive and big procedure. Important to make it easy to use and to understand, as you have tried to (FF)	Agree
	I think this is a useful framework for most decisions. It may be useful to add in an additional step for when comparisons need to be made between different policy options. (LH)	Agree. See above
	The real problems are not with the framework itself but with how it might be used. See comments below. I wonder if some hot tips to avoid particular issues might be useful (LV)	See above.
	Severity and N should be combined into one, e.g. using DALYs (MH)	
	The policy-makers I deal with are showing increasing interest in policy-modelling tools that can be used to compare the costs of alternative strategies, levels of investment, etc. The more flexible and interactive such tools are, the more	Agree about making the framework interactive and we will have an electronic version that is. Use of models might either feed into the framework or be a step after using the framework.

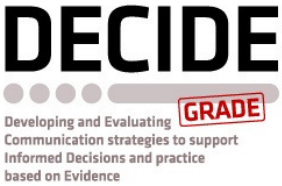
	Comments	Response
	likely they are to engage policy-makers in using the evidence. (PC)	Intention is to support and improve the use of logic and analysis.
	Seems to assume that the decision making process is more logic and analytic thatn often the case. Weak on the interaction between varies type of interests and how to handle (S)	Ask for examples of other comparable aids.
	I would like have seen why this is better (or its added value, which I am convinced it has) over other current decision-aids methods (XBC)	
Anything else **16.** Please include any other comments you have regarding the framework.	It may be useful to include a section asking whether it will be easy to measure whether the policy change has been successful, and if so, how that will be measured. (AP)	Agree. We have added monitoring at the end
	It may also be useful to prompt user to consider whether the evidence base is applicable for the context in which it will be used (e.g., are the systems in which similar policies were evaluated similar enough to the context in which it will be implemented in this case?) (AP)	Clarify that this is part of quality of evidence
	Possibly expand or add more on the implementation considerations? (ML)	Agree. See above
	General comment: it was a bit difficult to understand the logic and order of the decision tree in the mail and the levels of the different elements of the framework. You ask for comments on strategies for communicating evidence and to use this form that is linked to the framework for the whole process, which definitely is about more than communication.. But after having read a couple of times and tried it, I think	Thanks. Will add / improve intro.

	Comments	Response
	I understand.. and hopefully this feedback gives some sense.	
	This framework is a good tool, but as I see it, not generic for all situations or settings. The evidence for the first two steps (seriousness and number of people) and the last three-, maybe five steps (resources, value, impact, implementability and acceptability) will always be of another kind (and often lacking) than the evidence for intervention effects. This is where the GRADE instrument so far does not fit. Maybe this tool could help develop these areas of uncertainty in a decision making process further?	Clarify that GRADE approach to QoE does not apply to all types of evidence.
	Last comment; how will you deal with the questions of discounting, immediate costs to get future benefits? – or the long term costs of not investing now? (FF)	Add this to explanations for cost-effectiveness.
	Serious of the problem: This is very disease orientated and a service priority might also be helpful	Add guidance that will clarify this. Consider modifying the criterion/label.
	Resource use: Why is the appropriate perspective a government? For some countries governments have a comparatively narrow role. It might be useful to refer to criteria like stewardship, which from the WHO perspective would suggest that a government has an overarching role to ensure the quality of the whole health system	May not be. We will edit this.
	Value for money: I think you need to be more precise about what you mean by a societal perspective. A true societal perspective is very difficult to take	Agree. We will edit this.
	Impact on equity. The focus is on health inequalities. In many health care systems the focus is on other notions of equity e.g. equity of access for equal need	Agree this needs clarification

	Comments	Response
	Implementability (a bit of a made up/clumsy word?) What about how whether the change could be brought about in a timely fashion	Agree. We will look for a better word.
	Acceptability. Perhaps change nonmaleficence to plainer English? To what extent does justice overlap with equity?	OK.
	Example 1. How do you know if the utility values quoted are any good/appropriate to the context? It may help to provide a quality score as well	We have edited that.
	Example 1. The price year needs be stated for costs. Also I don't know whether the costs differences are relatively small, as a reader I can't tell. I also think that the unit of presentation (per patient) is misleading. The cost per patient may be small but the implications for the service may be large if there are many patients	OK. We need to consider these points.
	Example 1. I have grave concerns about the presentation of cost per QALY data. Normally the size of a negative ICER does not have a meaning as ICERs are discontinuous at 0. I think statements of 88% of simulations is too technical and also spurious precision. It may be best to turn it around and say there is 1 in 10 chance that it is not cost-effective	Not sure about this. Needs discussion.
	Example 1 – under the section on whether the option is feasible I think clarity about what unit or service you are talking about (LV)	Agree.
	I'm sure you will do it, but I'd love to see the methods for producing this as the degree to which it is vetted is an important consideration. That said, I'm anxious to try this out with state policy makers in the US. (MG)	Follow-up with MG.
	The main issue will be how the framework is implemented, i.e. how can It be incorporated into existing systems/structures. To be most effective it will need to be legislated or supported with strong	Agree.

	Comments	Response
	policies. This will also depend on the policy context. Other issues will be how it is supported with research evidence and who should make (or contribute to) the final decision (MH)	
	It is a good start, but I hope it can be made more flexible and interactive, so that users can quickly compare a range of options. (PC)	See above.
	Overall, I have found the framework a really interesting starting point, and the criteria in which it is articulated surely come up from a sensible process. I just have a few comments concerning its "right place" in the overall context of decision making at the system level. In a kind of schematic and simplistic way, I guess there are at least three different stages in the policy process: a) the agenda setting, where one has to decide about the issues to be addressed and their relative priority; b) the policy development/adoption, when actions to be taken are decided; c) policy evaluation, when the policy adopted are, more or less formally, assessed as for their impact. Focusing just on a) and b) stages, one of the major differences I see is that a) is to a great extent likely to be influenced by a number of sources, rather than just by the research information already available, including the claiming of the different stakeholders and information drawn at the system level (i.e. epidemiological as well as administrative and clinical data on health needs, on the characteristics of the supply structure and on its current activity, ect).	The focus is only on b. We agree.

	Comments	Response
	In its current version, the Decide framework seems to be focussed just on research information, thus it is likely to contribute just to the b) phase, policy adoption and, within that context, basically only for decisions concerning the adoption of individual health interventions/programmes. Another comment concerns a possible side effect of using the framework, which could be represented by the introduction of a bias towards areas/issues for which research information is available. Given that research is not equally distributed, but largely distorted and thus not addressing relevant issues and needs, it could be that those remain orphans if policy concentrate its efforts only where we do already have enough research information to act. This problem could be mitigated if the framework is proposed ì for supporting priority setting not across different problems/issues, but within: where you might have different policy options to address a problems, you could use the framework to get to the “best” option. In general, I really interesting efforts, surely deserving to proceed to next stages of development, including (but I am pretty sure it is already in your plans) to test its usefulness for policy makers and stakeholders. (RG)	This might be a risk, but seems unlikely.



Appendix 3

Evidence to Decision framework

Evidence to decision framework

[Question]?

Problem: [Problem]
Option: [Option]
Comparison: [Comparison]
Setting: [Setting]
Perspective: [Perspective]

Background: [Background]

Subgroup considerations: We have considered the following subgroups in relation to the following criteria and decided that separate recommendations are not needed for these subgroups. Our detailed judgments are provided below.

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	**Is the problem a priority?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐ ***Detailed judgements***	[Evidence]	"[Additional considerations]"

VALUES

CRITERIA: Is there important uncertainty or variability about how much people value the main outcomes?

JUDGEMENTS

☐ Important uncertainty or variability ☐ Possibly important uncertainty or variability ☐ Probably no important uncertainty or variability ☐ No important uncertainty or variability ☐ No known undesirable outcomes

Detailed judgements

RESEARCH EVIDENCE

The relative importance or values of the main outcomes of interest:

Outcome	Relative importance	Certainty of the evidence
[Outcome]	-	-
[Outcome]	-	-
[Outcome]	-	-
[Outcome]	-	-

ADDITIONAL CONSIDERATIONS

"[Additional considerations]"

BENEFITS & HARMS OF THE OPTIONS

CRITERIA: What is the overall certainty of the evidence of effectiveness?

JUDGEMENTS

☐ No included studies ☐ Very low ☐ Low ☐ Moderate ☐ High

CRITERIA: How substantial are the desirable anticipated effects?

JUDGEMENTS

☐ Don't know ☐ Not important ☐ Somewhat important ☐ Moderately important ☐ Very important ☐ ***Varies***

Detailed judgements

CRITERIA: How substantial are the undesirable anticipated effects?

JUDGEMENTS

☐ Don't know ☐ Very important ☐ Moderately important ☐ Somewhat important ☐ Not important ☐ ***Varies***

Detailed judgements

CRITERIA: Do the desirable effects outweigh the undesirable effects?

JUDGEMENTS

☐ No ☐ Probably No ☐ Don't know ☐ Probably Yes ☐ Yes ☐ ***Varies***

Detailed judgements

RESEARCH EVIDENCE

Summary of findings: *[Comparison]*

Outcome	Without [intervention] (per [#####]	With [intervention] (per [#####])	Difference (per [#####] (95%CI)	Relative effect (RR) (95%CI)	Certainty of the evidence (GRADE)
[Outcome]	[#####]	[#####]	**[#####] more/less**	**RR [#.##]** ([#.##] to [#.##])	[Rating]
[Outcome]	[#####]	[#####]	**[#####] more/less**	**RR [#.##]** ([#.##] to [#.##]	[Rating]
[Outcome]	[#####]	[#####]	**[#####] more/less**	**RR [#.##]** ([#.##] to [#.##]	[Rating]

Link to detailed evidence profile

Subgroup considerations:

Link(s) to summary of findings and judgments for subgroups

ADDITIONAL CONSIDERATIONS

"[Additional considerations]"

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
RESOURCE USE	How large are the resource requirements?	*Large costs* ☐ *Moderate costs* ☐ *Small* ☐ *Moderate savings* ☐ *Large savings* ☐ ***Varies*** ☐ **Detailed judgements**	[Evidence]	"[Additional considerations]"
RESOURCE USE	How large is the incremental cost relative to the net benefit?	*Very large ICER* ☐ *Large ICER* ☐ *Moderate ICER* ☐ *Small ICER* ☐ *Savings* ☐ ***Varies*** ☐ **Detailed judgements**	[Evidence]	"[Additional considerations]"
EQUITY	What would be the impact on health inequities?	*Increased* ☐ *Probably increased* ☐ *Uncertain* ☐ *Probably reduced* ☐ *Reduced* ☐ ***Varies*** ☐ **Detailed judgements**	[Evidence]	"[Additional considerations]"
ACCEPTABILITY	Is the option acceptable to key stakeholders?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐ **Detailed judgements**	[Evidence]	"[Additional considerations]"
FEASIBILITY	Is the option feasible to implement?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐ **Detailed judgements**	[Evidence]	"[Additional considerations]"

Overall judgement across all criteria	Undesirable consequences *clearly outweigh* desirable consequences in most settings ☐	Undesirable consequences *probably outweigh* desirable consequences in most settings ☐	The balance between desirable and undesirable consequences *is closely balanced or uncertain* ☐	Desirable consequences *probably outweigh* undesirable consequences in most settings ☐	Desirable consequences *clearly outweigh* undesirable consequences in most settings ☐
Type of recommendation	We recommend against the option or for the alternative ☐	We suggest not to use the option or to use the alternative ☐		We suggest using the option ☐	We recommend the option ☐
Justification	[Justification]				
	Detailed judgements				
Subgroup considerations	"[Subgroup considerations]"				
Implementation considerations	[Implementation considerations]				
Monitoring and evaluation considerations	[Monitoring and evaluation]				
Research priorities	[Research priorities]				

Evidence profile [title]

Author(s): [Authors]
Date: [YYYY-MM-DD]

[Insert GRADE evidence profile]

(Return)

References

(To make references appear here, place cursor in any text above this page and choose: Insert > Footnote…> Endnote > End of section)

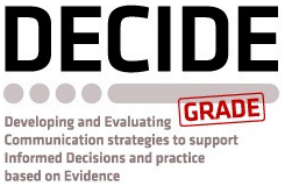
Definitions for ratings of the certainty of the evidence (GRADE)**

Ratings	Definitions
⨁⨁⨁⨁ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different* is low.
⨁⨁⨁◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[4] is moderate.
⨁⨁◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[4] is high.
⨁◯◯◯ Very low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[4] is very high.

*Substantially different: large enough difference that it might have an effect on a decision

**The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.

(Return)



Appendix 4

Evidence to Decision framework explanations

Evidence to decision framework

Explanations - Health system and public health recommendations

Purpose of the framework

The purpose of this framework is to help panels developing health system or public health guidelines move from evidence to recommendations. It is intended to:

- Inform panel members' judgements about the pros and cons of each option (intervention) that is considered
- Ensure that important factors that determine a recommendation (criteria) are considered
- Provide a concise summary of the best available research evidence to inform judgements about each criterion
- Help structure discussion and identify reasons for disagreements
- Make the basis for recommendations transparent to guideline users

Development of the framework

The framework is being developed as part of the DECIDE[1] project using an iterative process informed by the GRADE[2] approach for going from evidence to clinical recommendations, a review of relevant literature, brainstorming, feedback from stakeholders, application of the framework to examples, a survey of policymakers, user testing, and trials.

Description of the framework

The framework includes a ***table*** with the following columns:

- ***Criteria*** (factors that should be considered) for health system or public health recommendations
- ***Judgements*** that the panel members must make in relation to each criterion, which may include draft judgements suggested by the people who have prepared the framework
- ***Research evidence*** to inform each of those judgements, which may include links to more detailed summaries of the evidence
- ***Additional considerations*** to inform or justify each judgement

The framework also includes the following ***conclusions*** that the panel members must reach, which may include draft conclusions suggested by the people who have prepared the framework:

- The ***balance of consequences*** of the option being considered in relation to the alternative (comparison)
- The ***type of recommendation*** (against the option, for considering the option under specified conditions, or for the option)
- The ***recommendation*** in concise, clear and actionable text
- The ***justification*** for the recommendation, flowing from the judgements in relation to the criteria
- Any important ***subgroups considerations*** that may be relevant to guideline users
- Key ***implementation considerations*** (in addition to any that are specified in the recommendation), including strategies to address any concerns about the acceptability and feasibility of the option
- Suggestions for ***monitoring and evaluation*** if the option is implemented, including any important indicators that should be monitored and any needs for a pilot study or impact evaluation
- Any key ***research priorities*** to address important uncertainties in relation to any of the criteria

[1] Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE) is a 5-year project (running from January 2011 to 2015) co-funded by the European Commission under the Seventh Framework Programme. DECIDE's primary objective is to improve the dissemination of evidence-based recommendations by building on the work of the GRADE Working Group to develop and evaluate methods that address the targeted dissemination of guidelines.

[2] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.

Flexibility

The framework is flexible. Organisations may elect to modify the terminology (and language) that is used, the criteria, the response options and guidance for using the framework to ensure that the framework is fit for purpose.

Use of the framework

Suggestions for how to use the framework are provided in: *Framework for going from evidence to a recommendation – Guidance for health system and public health recommendations*, including suggestions for preparing frameworks, supporting use of the framework by guideline panels, and using the framework to support well-informed decisions by guideline users.

The final recommendation made by the guideline panel is a consensus based on the judgements of the panel members, informed by the evidence presented in the framework and the panel members' expertise and experience.

Explanations of the criteria in the framework

Why these criteria?

The criteria included in the framework are ones that have emerged from our literature review, brainstorming, feedback from stakeholders, application of the framework to examples, a survey of policymakers and user testing. It is possible that we will make further modifications based on continuing feedback, applications of the framework and user testing. Guideline developers may also want to make modifications, such as adding or removing criteria that are or are not important for them to consider. However, there is clear and consistent support for routinely including all of these criteria and, up to now, a lack of clear and consistent support for including other potential criteria.[1,2]

Detailed judgements

The judgements that need to be made are sometimes complex. Guideline panels are likely to find it helpful to make and record detailed judgements for some criteria using ***tables for detailed judgements***. This includes, for example, detailed judgements about the size of the effect for each outcome, the certainty of the evidence of the relative importance of the outcomes and resource use, and important subgroup considerations. Some criteria could be split into two or more separate criteria and some panels may elect to do this in order to highlight key considerations that are of particular importance for their guidelines. For example, there are several reasons why an option may not be acceptable to key stakeholders and these could potentially be considered as separate criteria.

From whose perspective?

Guideline panels should explicitly state the perspective that they are taking when making recommendations. This is especially important for determining which costs (resource use) to consider. It can also influence which outcomes and whose values are considered. For example, out-of-pocket costs are important from the perspective of an individual patient, whereas costs to the government are important from the perspective of the government. Health system and public health decisions are made on behalf of a population and a broad perspective is required. However, because of their mandate, some panels might take the perspective of the ministry of health or health department, whereas other panels might take a societal perspective (including all costs, regardless of who pays). Other perspectives (the distribution of the benefits, harms and costs) should be taken when considering the acceptability of the option to key stakeholders.

Large or small compared to what?

Some of the criteria imply a comparison; for example, the size of effects or resource requirements ***compared to what***? The comparisons or standards that are used are likely to be different for different organisations, guideline panels and jurisdictions. Some organisations or guideline panels may elect to specify the comparisons or standards that they will use. In the absence of such specified comparisons, guideline panel members should consider what their comparisons or standards are when they disagree, for example, about whether resource requirements are large. When the comparison being used is the source of their disagreement, they should agree on an appropriate comparison and include this as an additional consideration in the framework when it is relevant.

Guidance for making judgements

Suggestions for how to make judgements in relation to each criterion are provided in *Framework for going from evidence to a recommendation – Guidance for health system and public health recommendations*.

For each criterion there are four or five response options, from those that favour a recommendation against the option on the left to ones that favour a recommendation for the option on the right. In addition, most of the options include *varies* as a response option for situations when there is important variation across different settings for which the guidelines are intended and those differences are **substantial enough that they might lead to different recommendations for different settings.**

Questions to consider for each criterion and their relationship to a recommendation

For each criterion we suggest one or more detailed questions to consider when making a judgement and explain the relationship between the criterion and the recommendation.

CRITERIA	QUESTIONS	EXPLANATIONS
Is the problem a priority?	*Are the consequences of the problem serious (i.e. severe or important in terms of the potential benefits or savings)? Is the problem urgent? Is it a recognised priority (e.g. based on a national health plan)? Are a large number of people affected by the problem?*	The more serious a problem is, the more likely it is that an option that addresses the problem should be a priority (e.g., diseases that are fatal or disabling are likely to be a higher priority than diseases that only cause minor distress). The more people who are affected, the more likely it is that an option that addresses the problem should be a priority.
Is there important uncertainty about how much people value the main outcomes?	*How much do those affected by the option value each of the outcomes in relation to the other outcomes (i.e. what is the relative importance of the outcomes)? Is there evidence to support those value judgements, or is there evidence of variability in those values that is large enough to lead to different decisions?*	The more likely it is that differences in values would lead to different decisions, the less likely it is that there will be a consensus that an option is a priority (or the more important it is likely to be to obtain evidence of the values of those affected by the option). Values in this context refer to the relative importance of the outcomes of interest (how much people value each of those outcomes). These values are sometimes called 'utility values'.
What is the overall certainty of the evidence of effectiveness? [3]	*What is the overall certainty of this evidence of effects, across all of the outcomes that are critical to making a decision?*	The less certain the evidence is for critical outcomes (those that are driving a recommendation), the less likely that an option should be recommended (or the more important it is likely to be to conduct a pilot study or impact evaluation, if it is recommended).
How substantial are the desirable anticipated effects?	*How substantial (large)are the desirable anticipated effects (including health and other benefits) of the option (taking into account the severity or importance of the desirable consequences and the number of people affected)?*	The larger the benefit, the more likely it is that an option should be recommended.

[3] The "certainty of the evidence" is an assessment the likelihood that the effect will be substantially different from what the research found.

CRITERIA	QUESTIONS	EXPLANATIONS
How substantial are the undesirable anticipated effects?	*How substantial (large) are the undesirable anticipated effects (including harms to health and other harms) of the option (taking into account the severity or importance of the adverse effects and the number of people affected)?*	The greater the harm, the less likely it is that an option should be recommended.
Do the desirable effects outweigh the undesirable effects?	*Are the desirable effects large relative to the undesirable effects?*	The larger the desirable effects in relation to the undesirable effects, taking into account the values of those affected (i.e. the relative value they attach to the desirable and undesirable outcomes) the more likely it is that an option should be recommended.
How large are the resource requirements?	*How large an investment of resources would the option require or save?*	The greater the cost, the less likely it is that an option should be a priority. Conversely, the greater the savings, the more likely it is that an option should be a priority.
How large is the incremental cost relative to the net benefit?	*Is the cost small relative to the net benefits (benefits minus harms)?*	The greater the cost per unit of benefit, the less likely it is that an option should be a priority.
What would be the impact on health inequities?	*Would the option reduce or increase health inequities?*	Policies or programmes that reduce inequities are more likely to be a priority than ones that do not (or ones that increase inequities).

CRITERIA	QUESTIONS	EXPLANATIONS
Is the option acceptable to key stakeholders?	*Are key stakeholders likely to find the option acceptable (given the relative importance they attach to the desirable and undesirable consequences of the option; the timing of the benefits, harms and costs; and their moral values)?*	The less acceptable an option is to key stakeholders, the less likely it is that it should be recommended, or if it is recommended, the more likely it is that the recommendation should include an implementation strategy to address concerns about acceptability. Acceptability might reflect who benefits (or is harmed) and who pays (or saves); and when the benefits, adverse effects, and costs occur (and the discount rates of key stakeholders; e.g. politicians may have a high discount rate for anything that occurs beyond the next election). Unacceptability may be due to some stakeholders: • Not accepting the distribution of the benefits, harms and costs • Not accepting costs or undesirable effects in the short term for desirable effects (benefits) in the future • Attaching more value (relative importance) to the undesirable consequences than to the desirable consequences or costs of an option (because of how they might be affected personally or because of their perceptions of the relative importance of consequences for others) • Morally disapproving (i.e. in relationship to ethical principles such as autonomy, nonmaleficence, beneficence or justice)
Is the option feasible to implement?	*Can the option be accomplished or brought about?*	The less feasible (capable of being accomplished or brought about) an option is, the less likely it is that it should be recommended (i.e. the more barriers there are that would be difficult to overcome).

Explanations of the conclusions in the framework

Suggestions for how to make judgements in relation to each conclusion are provided in: *Framework for going from evidence to a recommendation – Guidance for health system and public health recommendations*. For each conclusion, we suggest one or more questions to consider when making a judgement and explain what is needed.

TERM	QUESTION	EXPLANATION
Overall judgement across all criteria	*What is the overall balance between all the desirable and undesirable consequences?*	An overall judgement whether the desirable consequences outweigh the undesirable consequences, or vice versa (based on all the research evidence and additional information considered in relation to all the criteria). Consequences include health and other benefits, adverse effects and other harms, resource use, and impacts on equity
Type of recommendation	*Based on the balance of the consequences in relation to all of the criteria in the framework, what is your recommendation?*	A recommendation based on the balance of consequences and your judgements in relation to all of the criteria, for example: • Not to implement the option • To consider the option only in the context of rigorous research • To consider the option only with specified monitoring and evaluation • To consider the option only in specified contexts • To implement the option
Recommendation (text)	*What is your recommendation in plain language?*	A concise, clear and actionable recommendation
Justification	*What is the justification for the recommendation, based on the criteria in the framework that drove the recommendation?*	A concise summary of the reasoning underlying the recommendation
Subgroup considerations	*What, if any, subgroups were considered and what, if any, specific factors (based on the criteria in the framework) should be considered in relation to those subgroups when implementing the option?*	A concise summary of the subgroups that were considered and any modifications of the recommendation in relation to any of those subgroups
Implementation considerations	*What should be considered when implementing the option, including strategies to address concerns about acceptability and feasibility?*	Key considerations, including strategies to address concerns about acceptability and feasibility, when implementing the option
Monitoring and evaluation considerations	*What indicators should be monitored? Is there a need to evaluate the impacts of the option, either in a pilot study or an impact evaluation carried out alongside or before full implementation of the option?*	Any important indicators that should be monitored if the option is implemented

TERM	QUESTION	EXPLANATION
Research priorities	*Are there any important uncertainties in relation to any of the criteria that are a priority for further research?*	Any research priorities

Explanations of terms used in summaries of findings

TERM	EXPLANATION
Outcomes	These are all the ***outcomes*** (potential benefits or harms) that are considered to be ***important*** to those affected by the intervention, and which are important to making a recommendation or decision. Consultation with those affected by an intervention (such as patients and their carers) or other members of the public may be used to select the ***important outcomes***. A review of the literature may also be carried out to inform the selection of the important outcomes. The importance (or value) of each outcome in relation to the other outcomes should also be considered. This is the ***relative importance of the outcome.***
95% Confidence Interval (CI)	A ***confidence interval*** is a range around an estimate that conveys how precise the estimate is. The confidence interval is a guide to how sure we can be about the quantity we are interested in. The narrower the range between the two numbers, the more confident we can be about what the true value is; the wider the range, the less sure we can be. The width of the confidence interval reflects the extent to which chance may be responsible for the observed estimate (with a wider interval reflecting more chance). ***95% Confidence Interval (CI)*** means that we can be 95 percent confident that the true size of effect is between the lower and upper confidence limit. Conversely, there is a 5 percent chance that the true effect is outside of this range.
Relative Effect or RR (Risk Ratio)	Here the ***relative effect*** is expressed as a ***risk ratio (RR).*** Risk is the probability of an outcome occurring. A ***risk ratio*** is the *ratio* between the risk in the intervention group and the risk in the control group. For example, if the risk in the intervention group is 1% (10 per 1000) and the risk in the control group is 10% (100 per 1000), the relative effect is 10/100 or 0.10. If the RR is exactly 1.0, this means that there is no difference between the occurrence of the outcome in the intervention and the control group. If the RR is greater than 1.0, the intervention increases the risk of the outcome. If it is a good outcome (for example, the birth of a healthy baby), a RR greater than 1.0 indicates a desirable effect for the intervention. Whereas, if the outcome is bad (for example, death) a RR greater than 1.0 would indicate an undesirable effect. If the RR is less than 1.0, the intervention decreases the risk of the outcome. This indicates a desirable effect, if it is a bad outcome (for example, death) and an undesirable effect if it is a good outcome (for example, birth of a healthy baby).

TERM	EXPLANATION
Certainty of the evidence (GRADE)[2] *Return to criterion*	The ***certainty of the evidence*** is an assessment of how good an indication the research provides of the likely effect; i.e. the likelihood that the effect will be substantially different from what the research found. By ***substantially different*** we mean a large enough difference that it might affect a decision. This assessment is based on an overall assessment of reasons for there being more or less certainty using the ***GRADE*** approach. In the context of decisions, these considerations include the applicability of the evidence in a specific context. Other terms may be used synonymously with *certainty of the evidence*, including ***quality of the evidence***, ***confidence in the estimate***, and ***strength of the evidence***. Definitions of the categories used to rate the certainty of the evidence (***high***, ***moderate***, ***low***, and ***very low***) are provided in the table below.

Definitions for ratings of the certainty of the evidence

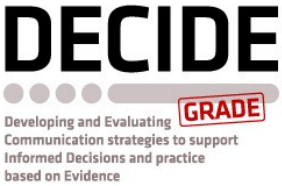
RATINGS	DEFINITIONS
⊕⊕⊕⊕ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be ***substantially different***[4] is low.
⊕⊕⊕◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[7] is moderate.
⊕⊕◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[7] is high.
⊕◯◯◯ Very Low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[7] is very high.

References

[1] Vogel JP, Oxman AD, Glenton C, Rosenbaum S, Lewin S, Gülmezoglu AM, Souza JP. Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: An international survey. Health Research Policy and Systems 2013; 11:19 DOI: 10.1186/1478-4505-11-19.

[2] Guindo LA, Wagner M, Baltussen R, Rindress D, Van Til J, Kind P, Goetghebeur MM: From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decision-making. Cost Eff Resour Alloc 2012, 10:9.

[4] Substantially different = large enough difference that it might have an effect on a decision



Appendix 5

Evidence to Decision framework example

Prepared by: Andy Oxman
Date: 26 November 2013

Evidence to decision framework – Public health guideline

Should the public health service encourage women age 50 to 69 to have screening mammography?

Problem: Breast cancer in women age 50 to 69
Intervention: Screening mammography
Comparison: No screening mammography
Setting: Sweden
Perspective: Government

Background: It is controversial whether the benefits of screening mammography outweigh the harms. The benefits of breast cancer screening must be balanced against the harms and the costs to the individual and to the healthcare system. The current Norwegian Breast Cancer Screening Programme consists of biennial two-view mammography screening offered to all women between 50 and 69 years. Mammography can identify asymptomatic breast cancer and breast cancer can be more effectively treated at the asymptomatic stage. Women with positive findings on mammography are offered further diagnostic tests, including imaging and biopsy.

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	**Is the problem a priority?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐ ***Detailed judgements***	Assume that the cumulative risk of invasive breast cancer for women age 50 to 69 without screening is about 44 per 1000 women. The risk of dying from breast cancer without screening is about 6 per 1000.	

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
VALUES	Is there important uncertainty about how much people value the main outcomes?	*Important uncertainty or variability* ☐ *Possibly important uncertainty or variability* ☐ *Probably no important uncertainty or variability* ☐ *No important uncertainty or variability* ☐ *No known undesirable outcomes* ☐ **Detailed judgements**	***The relative importance or values of the main outcomes of interest:*** **Outcome** / **Relative importance** / **Certainty of the evidence** *Total deaths* - ; *Breast cancer deaths* - ; *Breast surgery** - ; *Recalled for biopsies* - ; *Burden of screening†* - ; *Overdiagnosis‡* - ; There are few studies of how much women value life after treatment of breast cancer or living with metastatic breast cancer. A small Australian study found a utility value of about 0.8 for life after treatment of breast cancer and a utility value of about 0.3 for life with metastatic cancer. * Additional lumpectomies and mastectomies † Time, psychological and financial impacts of attending mammography ‡ Invasive or noninvasive breast cancer detected by screening that would not have been identified clinically and would not have resulted in symptoms or death in a person's lifetime. Consequences include surgery, radiotherapy, endocrine therapy, anxiety, distress, and stigmatization.	Some well-informed women elect to be screened and some do not.

BENEFITS & HARMS OF THE OPTIONS

CRITERIA	JUDGEMENTS
What is the overall certainty of the evidence of effectiveness?	☐ No included studies ☐ Very low ☐ Low ☐ Moderate ☐ High
How substantial are the desirable anticipated effects?	☐ Don't know ☐ Not important ☐ Somewhat important ☐ Moderately important ☐ Very important ☐ Varies **Detailed judgements**
How substantial are the undesirable anticipated effects?	☐ Don't know ☐ Very important ☐ Moderately important ☐ Somewhat important ☐ Not important ☐ Varies **Detailed judgements**
Do the desirable effects outweigh the undesirable effects?	☐ No ☐ Probably No ☐ Uncertain ☐ Probably Yes ☐ Yes ☐ Varies ***Detailed judgements***

RESEARCH EVIDENCE

Summary of findings: *Screening mammography for women age 40 to 49*

Outcomes (after 11.4 years)	No screening (per 10,000)*	Screening (per 10,000)	Difference (per 10,000) (95% CI)	Relative effect (RR) (95%CI)	Certainty of the evidence (GRADE)
Total deaths	350	372	- (14 fewer to 62 more)	1.06 (0.96-1.18)	⨁⨁⨁◯ Moderate
Deaths from breast cance	64	50	**14 fewer** (21 to 6 fewer)	0.79 (0.68-0.90)	⨁⨁⨁◯ Moderate
*Breast surgery**	163	215	**52 more** (36 to 70 more)	1.31 (1.22-1.42)	⨁⨁⨁◯ Moderate
Recalled for at least one biopsy	-	1201	**1201 more**	-	⨁⨁◯◯ Low
Burden of screening†	-	10 000	**10 000 more**	-	⨁⨁⨁⨁ High
Overdiagnosis‡	-	-	**-**	-	⨁◯◯◯ Very low

* Additional lumpectomies and mastectomies

† Time, psychological and financial impacts of attending mammography

‡ Any invasive or noninvasive breast cancer detected by screening that would not have been identified clinically and would not have resulted in symptoms or death in a person's lifetime. Consequences include surgery, radiotherapy, endocrine therapy, anxiety, distress, and stigmatization. Overdiagnosis is difficult to ascertain. Estimates range from 1·7% to 54% for women aged 50–59 years, and 7% to 21% for women aged 60–69 years.

Link to detailed evidence profile

Subgroup considerations:
Link(s) to summary of findings and judgments for subgroups

ADDITIONAL CONSIDERATIONS

Subgroup considerations:

High risk groups, including women with category 3 or 4 breast density, a previous breast biopsy, or a family history of breast cancer

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
RESOURCE USE	**How large are the resource requirements?**	*Large costs* ☐ *Moderate costs* ☐ *Small* ☐ *Moderate savings* ☐ *Large savings* ☐ ***Varies*** ☐	No evidence	The cost of a mammogram is about 650 kr. For 1,000,000 women screened biennially the cost would be around 325 million kr annually for mammograms. The full cost (including follow-up investigations and costs and savings from treatment) is not available.
	How large is the incremental cost relative to the net benefit?	*Very large ICER* ☐ *Large ICER* ☐ *Moderate ICER* ☐ *Small ICER* ☐ *Savings* ☐ ***Varies*** ☐ **Detailed judgements**	The estimated cost per QALY gained is about 540 000 kr for screening every 2 years. The cost per QALY gained is less for high risk groups. For example, biennial mammography costs less than 300,000 NOK per QALY gained for women with either a previous breast biopsy or a family history of breast cancer.	**Subgroup considerations**: High risk groups, including women with category 3 or 4 breast density, a previous breast biopsy, or a family history of breast cancer

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
EQUITY	**What would be the impact on health inequities?**	*Increased* ☐ *Probably increased* ☐ *Uncertain* ☐ *Probably reduced* ☐ *Reduced* ☐ ***Varies*** ☐		Not offering screening mammography in the public health service might increase inequities for women who want to be screened.
ACCEPTABILITY	**Is the option acceptable to key stakeholders?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐	Qualitative studies suggest that women value screening mammography due to their perception that it reduces mortality. Few women have considered the harms of screening or further investigation. Studies show a wide variation in women's preferences, with the risk of false positives decreasing preferences for screening in some cases, while in others there is a willingness to be recalled for further investigation if it might increase the chance of detecting cancer earlier. Some women do not attend for screening because of fear, fatalistic beliefs, absence of symptoms, or work or family responsibilities. Most women prefer to share mammography decision making with their doctor.	Screening mammography is advocated by some people and organizations.
FEASIBILITY	**Is the option feasible to implement?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☐		Waiting lists for both screening and diagnostic mammography?

Overall judgement across all criteria	Undesirable consequences *clearly outweigh* desirable consequences in most settings ☐	Undesirable consequences *probably outweigh* desirable consequences in most settings ☐	The balance between desirable and undesirable consequences *is closely balanced or uncertain* ☐	Desirable consequences *probably outweigh* undesirable consequences in most settings ☐	Desirable consequences *clearly outweigh* undesirable consequences in most settings ☐
Type of recommendation	We recommend not offering the intervention ☐	We suggest not offering the intervention ☐	We suggest offering the intervention ☐	We recommend offering the intervention ☐	
Recommendation (text)	[Recommendation]				
Justification	[Justification]				
Subgroup considerations	[Subgroup considerations]				
Implementation considerations	[Implementation considerations]				
Monitoring and evaluation considerations	[Monitoring and evaluation]				
Research priorities	[Research priorities]				

Evidence profile [title]

Author(s): [Authors]
Date: [YYYY-MM-DD]

[Insert GRADE evidence profile]

(Return)

References

1. Redman S, Spencer EA, Sanson-Fisher RW. The role of mass media in changing health-related behaviour: a critical appraisal of two models. Health Promotion International 1990;5:85-101
2. Brinn MP, Carson KV, Esterman AJ, Chang AB, Smith BJ, Sowden AJ. Mass media interventions for preventing smoking in young people. Cochrane Database of Systematic Reviews 2010, Issue 11
3. Wammes A, Van Leeuwen, Lokman S. Evaluatierapport DVP-Campagne 2006 'Je bent niet gek als jeniet blowt'. Utrecht: Trimbos Institute, 2007
4. Rossi PH, Freeman HE. Evaluation: A Systematic Approach. 7th edition. Newbury Park: Sage Publications, Inc, 2003.
5. Hibell B, Guttormsson U, Ahlström S, Balakireva O, Bjarnason T, Kokkevi A, et al. The 2011 ESPAD Report - Substance Use Among Students in 36 European Countries. Stockholm: Swedish Council for Information on Alcohol and other Drugs, European Monitoring Centre for Drugs and Drug Addiction, 2012.
6. European Monitoring Centre for Drugs and Drug Addiction. The State of the Drugs Problem in Europe. Luxembourg: Publications Office of the European Union, 2012
7. Leshner AI. Drug abuse and addiction treatment research: the next generation. Archives of General Psychiatry 199754:691-4.
8. Leshner AI. Science-based views of drug addiction and its treatment. JAMA 1999;282(14):1314-6.
9. Camí J,¬Farré M. Drug addiction. New England Journal of Medicine 2003;348(10):975-86
10. McLelland T, Lewis DC, O'Brien CP, Kleber HD. Drug dependence, a chronic medical disease: Implications for treatment, insurance and outcome evaluation. JAMA 2000;284:1689-95
11. Swendsen J, Conway KP, Degenhardt L, Dierker L, Glantz M, Jin R, et al. Socio-demographic risk factors for alcohol and drug dependence: the 10-year follow-up of the national comorbidity survey. Addiction 2009;104(8):1346-55
12. Ashton B. The American Star comes to England. Drug and Alcohol Findings 2003;8:21-6
13. Leshner AI. Drug abuse and addiction treatment research: the next generation. Archives of General Psychiatry 1997; 54:691-4.
14. Nestler EJ, Aghajanian GK. Molecular and cellular basis of addiction. Science 1997;278:58-63
15. Wise RA. Drug-activation of brain reward pathways. Drug and Alcohol Dependence 1998;51:13-22
16. Scheier LM, Grenard JL. Influence of a nationwide social marketing campaign on adolescent drug use. Journal of Health Communication 2010;15(3):240-71
17. Hornik R. Personal influence and the effects of the National Youth Anti-Drug Media Campaign. The ANNALS of the American Academy of Political and Social Science 2006;608(1):282-300
18. Colorado Meth: Use & Attitudes Survey 2011. The Meth Project 2011
19. Georgia Meth: Use & Attitudes Survey 2011. The Meth Project 2011
20. Hawaii Meth: Use & Attitudes Survey 2011. The Meth Project 2011
21. Idaho Meth: Use & Attitudes Survey 2010. The Meth Project 2010
22. Wyoming Meth: Use & Attitudes Survey 2011. The Meth Project 2011
23. Fishbein M, Hall-Jamieson K, Zimmer E, von Haeften I, Nabi R. Avoiding the boomerang: testing the relative effectiveness of antidrug public service announcements before a national campaign. American Journal of Public Health 2002;92(2):238-45
24. Yzer MC, Cappella JN, Fishbein M, Hornik R, Ahern RK. The effectiveness of gateway communications in anti-marijuana campaigns. Journal of Health Communication 2003;8(2):129-43
25. Palmgreen P, Donohew L, Lorch EP, Hoyle RH, Stephenson MT. Television campaigns and adolescent marijuana use: tests of sensation seeking targeting. American Journal of Public Health 2001;91(2):292-6.

Definitions for ratings of the certainty of the evidence (GRADE)**

Ratings	Definitions
⨁⨁⨁⨁ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different* is low.
⨁⨁⨁◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[4] is moderate.
⨁⨁◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[4] is high.
⨁◯◯◯ Very low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[4] is very high.

*Substantially different: large enough difference that it might have an effect on a decision

**The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.

(Return)