Rosenbaum, S. E., C. Glenton & J. Cracknell. (2008). User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library. *Bmc Medical Informatics and Decision Making*, 8.

BMC Medical Informatics and Decision Making

BioMed Central

Research article

User experiences of evidence-based online resources for health professionals: User testing of *The Cochrane Library*

Sarah E Rosenbaum*†[1], Claire Glenton†[1] and Jane Cracknell[2]

Address: [1]Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs plass, N-0130 Oslo, Norway and [2]Rigshospitalet, Dep. 33-4-52, Blegdamsvej 9, 2100 Copenhagen, Denmark

Email: Sarah E Rosenbaum* - sarah@rosenbaum.no; Claire Glenton - claire.glenton@kunnskapssenteret.no; Jane Cracknell - jane_cracknell@yahoo.com

* Corresponding author †Equal contributors

Published: 28 July 2008

BMC Medical Informatics and Decision Making 2008, **8**:34 doi:10.1186/1472-6947-8-34

Received: 22 April 2008
Accepted: 28 July 2008

This article is available from: http://www.biomedcentral.com/1472-6947/8/34

Abstract

Background: Evidence-based decision making relies on easy access to trustworthy research results. *The Cochrane Library* is a key source of evidence about the effect of interventions and aims to "promote the accessibility of systematic reviews to anyone wanting to make a decision about health care". We explored how health professionals found, used and experienced The Library, looking at facets of user experience including findability, usability, usefulness, credibility, desirability and value.

Methods: We carried out 32 one-hour usability tests on participants from Norway and the UK. Participants both browsed freely and attempted to perform individually tailored tasks while "thinking aloud". Sessions were recorded and viewed in real time by researchers. Transcriptions and videos were reviewed by one researcher and one designer. Findings reported here reflect issues receiving a high degree of saturation and that we judge to be critical to the user experience of evidence-based web sites, based on principles for usability heuristics, web guidelines and evidence-based practice.

Results: Participants had much difficulty locating both the site and its contents. Non-native English speakers were at an extra disadvantage when retrieving relevant documents despite high levels of English-language skills. Many participants displayed feelings of ineptitude, alienation and frustration. Some made serious mistakes in correctly distinguishing between different information types, for instance reviews, review protocols, and individual studies. Although most expressed a high regard for the site's credibility, some later displayed a mistrust of the independence of the information. Others were overconfident, thinking everything on *The Cochrane Library* site shared the same level of quality approval.

Conclusion: Paradoxically, *The Cochrane Library*, established to support easy access to research evidence, has its own problems of accessibility. Health professionals' experiences of this and other evidence-based online resources can be improved by applying existing principles for web usability, prioritizing the development of simple search functionality, emitting "researcher" jargon, consistent marking of site ownership, and clear signposting of different document types and different content quality.

Background

The value of evidence-based medicine (EBM) – using updated, relevant and trustworthy evidence to inform medical decisions is widely acknowledged [1]. Recently the British Medical Journal nominated EBM as one of the 15 most important milestones in medicine since 1840 [2]. Easy access to high quality research has the potential to improve patient care, but there are obstacles that face health professionals attempting to use evidence in their practice. In an Australian survey, physicians identified insufficient time (74%), limited search skills (41%) and limited access to evidence (43%) as impediments to making better use of research data [3].

Systematic reviews directly address several of these barriers, as their summarized form reduces the amount of time and search skills needed to access and appraise many individual studies [4]. A systematic review is a summary of individual studies addressing a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise the relevant research, and to collect and analyse data from the included studies. The Cochrane Collaboration is an international organisation of volunteers dedicated to producing systematic reviews of rigorous methodological quality. These reviews are published in one of the databases on *The Cochrane Library* [5], a web site that has the potential to further simplify the task of finding trustworthy evidence. Additionally the Library hosts other databases for systematic reviews, health technology assessments and randomized controlled trials, making it a central online collection of varying types of evidence from a variety of sources.

Part of the mission of The Cochrane Collaboration is "to promote the accessibility of systematic reviews to anyone wanting to make a decision about health care". The organization also aims to produce reviews that are easy to read and understand by someone with a basic sense of the topic [6]. But does the Library web site support the Collaboration's goals of clarity and ease of use, as well as the overreaching mission of making evidence accessible for decision making? We wanted to explore this question through observing the experiences of health professionals using *The Cochrane Library*. We were interested not only in site-specific problems but also in issues that might be relevant to other web sites publishing collections of evidence-based content.

User experience

Usability testing is a method that is widely used in the field of web design to uncover errors and areas of improvement by observing users solving given tasks on the site [7,8]. There is increased recognition of the limitations of examining only task-related problems when attempting to understand why users' interactions with web sites might succeed or fail. Attention to the user's whole experience has begun to gain ground in the field of human-computer interaction [9]. Morville's "honeycomb" model (see Figure 1) distinguishes between seven separate facets of user experience, including findability, accessibility, usability, usefulness, credibility, desirability and value [10].

A brief explanation of these terms:

Findability: can users locate what they are looking for?

Accessibility: are there physical barriers to actually gaining access, also for people with handicaps?

Usability: how easy and satisfying is this product to use?

Usefulness: does this product have practical value for this user?

Credibility: is it trustworthy?

Desirability: is it something the user wants? Has a positive emotional response to?

Value: does this product advance the mission of the organization behind it?

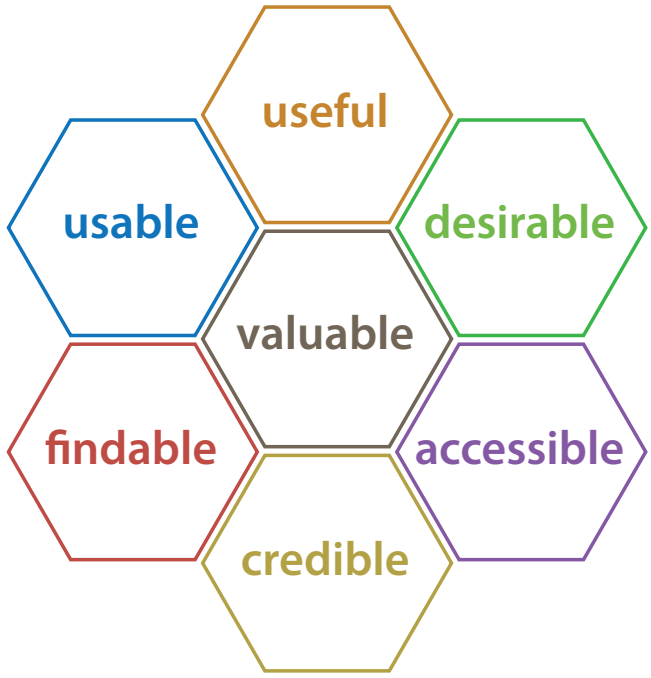


Figure 1
The honeycomb model of user experience, reproduced here with permission from Peter Morville, Sematic Studios LLC.

Our study aimed to explore the user experience of health professionals trying to find evidence in *The Cochrane Library*, building on methods from usability testing. In this article we use the honeycomb model to organize findings from our study to illustrate more general potential pitfalls and challenges particular to evidence-based online resources. At the end we suggest some guidelines for designers, writers and developers working to improve the user experience of these types of sites.

Methods

We carried out two series of user tests in 2005 (Test 1) and 2006 (Test 2), with participants from Norway and UK. The publisher of the site, Wiley-Blackwell, made changes to the site after Test 1, partly based on the results we uncovered. Most of these changes regarded branding at the top of the site, making *The Cochrane Library* the prominent identity and toning down the logo and universal navigation of the publisher. Therefore we altered the interview guide of Test 2 in small ways so that the questions would match the changes that had been made. See Additional file 1 for the complete interview guide we used in Test 2.

We limited our selection to health professionals who used the Internet and had some knowledge of systematic reviews, to ensure that the results of the interface testing would not be confounded by unfamiliarity with the media or the site's content. We sent email invitations to lists of previous attendees of evidence-based practice workshops, employees in the Directorate of Health and Social Affairs in Oslo and individuals in evidence-based health care networks in Oxford. Volunteers who responded were screened by phone or email to assess whether they fitted the requirements, and also to find relevant topics of interest so that we could individually tailor test questions. We also asked them about their online searching habits, and what sources of online information they usually used in connection with work. We did not reveal the name of the site we were testing during recruitment. Test persons were promised a gift certificate worth the equivalent of $80 USD or a USB memory stick if they showed up for the test.

Tests were performed individually and took approximately one hour. The test participant sat at a computer in a closed office together with the test leader who followed a semi-structured test guide. We recorded all movement on the computer desktop through use of Morae usability test software [11] and video-filmed the participant, who was prompted to think out loud during the whole session. We projected the filming of the desktop and the participant as well as the sound track, to another room where two observers transcribed, discussed, and took notes.

The data was anonymous to the degree that participants' names were not connected to video, audio or text results. We received written permission to store the recordings for five years before deleting it, guaranteeing that video/audio tapes would not be used for any purpose outside of the study and not be published/stored in places of public access. The protocol was approved by the Norwegian Social Science Data Services and found in line with national laws for privacy rights.

We began the test with preliminary questions about the participant's profession, use of Internet, and knowledge of *The Cochrane Library*. We then asked the participant to find specific material published on the Library starting from an empty browser window. Once on the site, we asked about their initial reactions to the front page, and they were invited to browse freely, looking for content of interest to themselves. Then we asked them to perform a series of tasks, some of which involved looking for specific content about topics tailored to their field or professional interests. For instance, a midwife was asked to find:

- all information on the whole library that dealt with prevention of spontaneous abortion

- a specific review about the effect of caesarean section for non-medical reasons

- all new Cochrane Reviews relevant to the topic "music used to relieve pain".

Other general tasks included finding help, finding the home page, and finding information about Cochrane. We also had specific tasks leading to searching and to reading a review. At the end, we asked if they had any general comments to the site and suggestions to how it could be improved.

Our analysis was done in two phases. The aim of the first analysis was to provide the stakeholders and site developers with an overview and a prioritizing of the problems we had identified. At least two of us carried out content analysis of the transcripts, independently coding each test. These codes were then compared, discussed and merged. The topics were then rated according to the severity of the problem for the user. We rated severity in three categories: high (show-stopper, leads to critical errors or hinders task completion), medium (creates much frustration or slows user down), or low (minor or cosmetic problems).

The second analysis was done to lift more generalizable issues underlying this article out of the site-specific data. We re-sorted the findings into the seven user-experience categories from the honeycomb model by re-reading the

transcript, checking the context where the problems came from, and evaluating which of the seven categories best fit each finding. Severity-of-problem ratings from the first analysis were kept in the second analysis.

We did not evaluate accessibility (the degree to which the website complied with standards of universal accessibility, for instance as defined by the Web Accessibility Initiative [12]), since user testing methods are not an effective way of gathering data on various aspects of this issue.

The findings presented here are a selection of issues that received a high degree of saturation in our tests, and that we judge to be critical ("high severity") to the user experience of evidence-based web sites in general. This judgement is based on basic principles for web usability [7,13-15] as well as the principles underlying evidence-based health care: to successfully search for, critically appraise and apply evidence in medical practice [16].

Most of the findings here are still of relevance to *The Cochrane Library* in its current format, though we have included some observations of problems that are now resolved, because they illustrate issues that are potentially important for others. Our aim is not to write a critical review of the library, but to highlight issues we found that can be important to user experience of evidence-based web sites for health professionals.

Results

Participant profiles

We tested a total of 32 persons (See Table 1 for participant details). Test 1 included 13 persons from Norway, and Test 2 included five persons from Norway and 14 from the UK. Twenty-one of the 32 participants were non-native English speakers accustomed to reading in English.

Participants were educated in nursing/midwifery (10); medicine (8); dentistry (4); physiotherapy (4); social sci-

Table 1: Participant details

	Gender	Age	Profession	Internet use: Frequency	Native language	Place of residence
1	F	44	Midwife	Daily	Norwegian	Oslo
2	F	43	Sociologist, advisor in health-related govt. institution	Daily	Norwegian	Oslo
3	F	53	Physical therapist/teacher	1–2 times a month	Norwegian	Oslo
4	F	45	Midwife/researcher	Daily	Other (not English)	Oslo
5	F	-	advisor in health-related govt. institution	Up to 5 times a week	Norwegian	Oslo
6	F	-	Masters in nursing science, lectures at college level	Daily	Norwegian	Oslo
7	F	39	Midwife/teacher	Daily	Norwegian	Oslo
8	M	49	Medical Doctor/dept. director at health-related govt. institution	Daily	Norwegian	Oslo
9	F	28	Psychologist at health station for youth		Norwegian	Oslo
10	M	40–50	Medical Doctor/senior advisor at health-related govt. institution	Daily	Norwegian	Oslo
11	F	56	Sociologist/Masters in health admin./advisor at health-related govt. institution	Almost everyday	Norwegian	Oslo
12	M	25–35	Physical therapist	Daily	Norwegian	Oslo
13	F	28	Physical therapist at county health station	Up to 5 days a week	Norwegian	Oslo
14	M	43	Psychologist at hospital	Daily	Norwegian	Oslo
15	F	34	Medical Doctor at hospital	Up to 5 days a week	Norwegian	Oslo
16	M	49	Medical Doctor at hospital	Daily	Norwegian	Oslo
17	F	54	Midwife/teacher	3 times a week	Norwegian	Oslo
18	F	23	Nurse (recently graduated)	3 times a week	Norwegian	Oslo
19	F	42	Research nurse	5–10 hours a week	Danish	Oxford
20	F	-	Pediatric Nurse	10–20 hours a week	English	Oxford
21	F	45	Consultant, public health. Clinical dentist, doing an Mba	10–20 hours a week	English	Oxford
22	M	35	Medical Doctor	10–20 hours a week	English	Oxford
23	F	31	Psychiatrist	10–20 hours a week	English	Oxford
24	F	46	General practitioner	20–40 hours a week	English	Oxford
25	F	41	Mental Health nurse	5–10 hours a week	English	Oxford
26	M	66	Consultant Dentist Public Health	Less than 5 hours a week	English	Oxford
27	F	32	Nursing, Post-doc in nursing-related field	10–20 hours a week	English	Oxford
28	F	40	Clinical orthodontist	Up to 5 times a week	English	Oxford
29	F	45	Occupational therapist	Less than 5 times a week	Other (not English)	Oxford
30	F	50	Nursing, Midwife, starting Phd	Up to 5 times a week	English	Oxford
31	M	-	Dentist	Daily	English	Oxford
32	M	54	General practitioner	5–10 hours a week	English	Oxford

ences (3); psychology (2); and occupational therapy (1). They were currently working as health professionals in primary or secondary care (17); as government advisors working with health-related issues (7); as teachers at nursing/physiotherapy schools or universities (4); as research nurses (3); or as an editor for a patient information website (1).

Most used the Internet daily or several times a week, and much of this use was work-related. All had searched the Internet for health-related information or evidence. Most participants reported that they normally looked for information in response to a specific problem. A few of them had strategies to keep up to date within a certain field on a more regular basis. When in need of information, the most common sources mentioned were colleagues, research databases, and the Internet. All but one participant had some previous knowledge of The Cochrane Collaboration and 25 of the 32 participants could provide at least a basic description of the term "systematic review". Twenty-six said that they had visited *The Cochrane Library* site previously.

The findings that we included in this article are listed in Table 2.

Findability

Finding the website

Finding the site was an obstacle for the majority of participants in Test 1. Despite the fact that 11 of 13 of these participants said they had visited *The Cochrane Library* before, the same number were not able to find the site without considerable confusion, and six of these 11 did not find the site at all until they were helped by the test facilitator. Although most participants in Test 2 had more success, finding the site remained a problem for some. One of these, a EBM-skilled UK participant, used 23 minutes to arrive at *The Cochrane Library* from a blank browser page.

Table 2: Main findings, sorted into the facets of the honeycomb user experience model

Findability	Difficulty finding the web site through Google or other external search Difficulty finding specific content on the site, using on-site search - non-English participants spelled search queries wrong - search engine too sensitive - keywords search didn't work properly - simple search produced unexpected results (i.e.: too few or too many of wrong type) - search results were misinterpreted, users confused document types - confusion when retrieving only a small number of search results Topics navigation not used or not seen Minimum of browsing even when encouraged to look around the site
Usability	Unfamiliar language/jargon caused confusion Text too small Too dense, too much text (front page, Help, More information pages) Important content too far down on page (review pages) Not interested in reading whole review Forrest plots unfamiliar and not intuitively located
Credibility	Users trusted content in The Cochrane Library Confusion about site ownership/neutrality due to dominance of publisher identity and universal navigation, weakens trust Misunderstanding about editorial quality evaluation – thinking all content on the whole site content has been reviewed by Cochrane
Usefulness	Assuming the library only dealt with medical topics (and not topics such as dentistry, nutrition, acupuncture) Misunderstanding targeted texts on front page, thinking content would be tailored for these groups Perceived as an academic resource Plain language summaries appreciated
Desirability	Site seemed off-putting, overwhelming Site can be alienating (research/academic identity and language)
Value	Felt Cochrane represented golden standard for systematic reviews Site is too difficult, would go elsewhere
Accessibility	Not evaluated

Much of this trouble stemmed from the participants' failed attempts to find Cochrane through Google search technology. These searches often failed because Google did not rank *The Cochrane Library* on the top of the first results page when queried for "Cochrane" or "Cochrane Library". In part this may be due to the fact that only the top few pages of *The Cochrane Library* were open to indexing in Google, affecting the ranking of the site. Several participants followed other links that appeared higher up on the results list, including links leading to the previous publisher of the site and to The Cochrane Collaboration site, expecting they would lead to the Library. After arriving at these other sites, participants continued to express confusion as to where they were because they found Cochrane-related content.

Problems searching for content

Finding specific content was also a major problem once participants arrived at *The Cochrane Library*. Participants attempted to solve most tasks by performing a search. Even when participants were asked to "take a look around the site", 75% started this task with a search. Few of our participants used the advanced search functionality. The simple search was the single most used feature in these tests, and many of these searches failed, leaving participants with a negative impression of the search functionality in the Library. Some participants compared *The Cochrane Library* to PubMed search, which they found easier to use.

Misspelling was the most common search-related mistake made by non-English participants. They were used to getting help with this from other search engines that was not provided by *The Cochrane Library* search: "*If I get the spelling wrong, Google will help*". Another problem this group experienced was recalling precise terms (for instance recalling "overweight" but not "obesity"). The publisher redesigned parts of the search interface after Test 1. However in Test 2 the non-native English participants still had considerable problems finding content, mainly due to problems with spelling and recall of correct terms.

Search results were often misinterpreted. One of the most critical problems we observed was participants' confusion regarding what they were finding. Many participants did not notice that hits occurred in different databases in *The Cochrane Library* and thought all hits were completed Cochrane Reviews. We observed participants clicking on and reading review protocols and reports of individual clinical trials, mistaking them for systematic reviews.

The search engine was also too sensitive. For instance "huntingtons" gave no hits, while "huntington's" did. "Keywords" option did not provide stabile results.

Participants were also confused when their searches produced few or no search results. Some misinterpreted getting few hits as being the result of a bad search. The concept underlying the Cochrane Database of Systematic Reviews of one review per subject did not seem apparent. In addition, non-native English speakers interpreted a lack of hits as a result of their own bad English even though this might not have been the case.

Problems browsing for content

Test persons did not browse much, though this may have had to do with their problems understanding the organisation of the site. Few people were able to describe how the content was structured by viewing the front page and nobody could point to a menu with any certainty. Only one test person used the "Topics" entry at the top of the front page, though it was not apparent whether other participants did not see it or preferred not to use it.

Usability

Language and jargon

Participants reacted to the use of jargon throughout the site. Some of the jargon was site-specific (such as the term "*record*" which led to full texts) and some was tied to research terminology (for instance "*protocol*"). The use of jargon gave the impression that the site was for academic use only and effectively discouraged participants from using several of the site's functions.

Legibility and layout

Most felt that there was too much text on the front page and that the type was too small. The participants that clicked on the "Help" and the "More Information" section also found them very dense.

> "*It's very messy. Do I have to read all of this?*"

There was lots of frustration about the screen being taken up by other things than the review text such as the top banner space. Several participants made negative comments about having to scroll down to see full front page.

> "*The actual content is stuck in this little area down here.*"

Reading pattern

We were interested in how participants read reviews and asked them to show us how they normally would approach document if they had limited time (two to five minutes). Most referred to the conclusion section. Several said they would read the abstract, while some mentioned the objectives, results, and background sections. Most said that they normally would not be interested in reading a whole review.

We asked participants specifically about the forest plot graphs in the Cochrane Reviews, as they present a lot of information in a summarized form that could be useful for a reader in a hurry. Some participants found them helpful; others found them confusing. They were very difficult to comprehend for those participants who had not seen them before, and were not intuitively located.

Credibility

When asked if they would trust the information on *The Cochrane Library*, all participants replied that they would, often because of a familiarity with the Cochrane name and more or less vague ideas about the quality of Cochrane products: "*because it's very respected*"; "*it's a reputable name*"; "*because I've heard good things about it.*"

In Test 1, however, we observed potential challenges to this trust because of confusion about site/content ownership. This was primarily tied to the prominence of the Library's publisher Wiley-Blackwell on the website. Wiley's logo was placed higher up on the page than Cochrane's, and Wiley's Home, About Us, Contact Us, and Help buttons were assumed to be Cochrane Library buttons by most participants. Participants who used these buttons often did not realise that they were no longer in *The Cochrane Library*. When asked to describe the relationship between Wiley and *The Cochrane Library*, many described *The Cochrane Library* as a sub-group of Wiley:

> "*It gives me sort of pharmaceutical industry associations. I think that The Cochrane Library is a subgroup (of Wiley).*"

Several changes were made to the website in order to address these issues after Test 1, and participants in Test 2 did not display the same confusion.

We also observed that *The Cochrane Library*'s perceived credibility could be over-interpreted. The only contents on *The Cochrane Library* that are "Cochrane approved" are the reviews listed in the Cochrane Database of Systematic Reviews. Despite this fact, some participants assumed that everything in the Library was "*Cochrane-approved*", including the trials, reviews and reports in the individual databases: "*This will just have things that Cochrane have looked at*"; "*If I was looking for a piece of evidence and I found it on Cochrane I would think that it was high quality*."

Usefulness

Some participants assumed that *The Cochrane Library* only dealt with medical topics and did not expect to find information on topics such as dentistry, nutrition, or acupuncture. The Library was also perceived by some as primarily an academic resource: "*I've tended to think that this is where researchers go to add to the body of knowledge or to see what there is, they'd use this (to build up) Clinical Evidence or Bandolier.... but if I was wanting to get back to the source of information, this is where I would want to go.*"

The website has attempted to signal that it is a resource for all types of healthcare decision-makers by adding buttons on the front page entitled "For Clinicians"; "For Researchers"; "For Patients"; and "For Policy makers". These lead to short descriptions of what *The Cochrane Library* can offer each of these groups. However, while some participants thought these were advertising because of their position in the right-hand column, several others assumed that they led to specially adapted versions of *The Cochrane Library*, and were disappointed when this turned out not be the case:

> "*I'm surprised that there's a link through to patients here. (...) I didn't realise that it was so well-developed along those lines.*"

> "*Oh, so it's an (advert)... I was hoping it would give me a tailored search programme, a bit like NLH, which asks you "are you a GP...*"

Others disliked these distinctions between different target groups: "*I don't know why clinicians should differ from researchers. We all need to have "high quality information at our fingertips.*"

Several participants were positive to the fact that patients' information needs were being addressed in the form of the Plain Language Summaries they found in the Cochrane Reviews. They saw these products as helpful both for communicating with patients and for understanding the research results themselves.

> "*I wouldn't want to go and read all the nitty gritty. The short bits, the one page was useful.*"

Desirability

Two thirds of the participants complained that the site looked messy and difficult to use, that there was too much information. All expressed frustration with failed attempts to find relevant content. Participants wanted a web site they could get into quickly, find what they were looking for, and get out again. "*Crowded*," "*busy*," "*cluttered*," "*a lot going on*," "*difficult to find any one particular thing*" were typical comments. Some participants felt "*overwhelmed*," "*bombarded*" and "*stupid.*"

While most expressed interest in this type of evidence-based resource, many were cautious, or concerned that they lack the necessary skills: A nurse commented: "*This is maybe more for doctors*." A physician who had trouble finding specific content chose to search for "*dementia*" during

a test task, and explained why: *"That's kind of how I'm feeling right now."*

Value

At the beginning of the test all participants said they expected to be able to find content that was relevant for them on *The Cochrane Library*. Most felt that Cochrane Reviews represented the golden standard for systematic reviews. Many were put off by the amount of information and concerned about the time it would take them to find what they were looking for.

> *"Not easy to get around"; "Most of us don't have time to get around"; "So many pages are better designed, so you just get fed up and frustrated and go somewhere else."*

Discussion

Our study shows that health professionals' experiences of *The Cochrane Library* were considerably less than optimal. Test participants had much difficulty locating both the site and the evidence. Non-native English speakers were at an extra disadvantage when retrieving relevant documents. Many participants displayed feelings of ineptitude, alienation and frustration. Some made serious mistakes in correctly identifying different information types. Although nearly all expressed a high regard for the credibility of The Cochrane Collaboration, some later displayed a mistrust of the independence of the information. Others were overconfident, thinking everything on *The Cochrane Library* site had been quality-approved through an editorial evaluation, transferring the quality association they had of Cochrane Reviews to the entire content of the library.

There are few published usability studies of health professionals using online health libraries or other similar collections of evidence-based medical literature. A commercial company carried out parallel testing of *The Cochrane Library* for Wiley-Blackwell in 2005 and 2006. Their unpublished reports showed findings that were by and large similar to ours, though included only participants living and working in the UK and therefore did not duplicate the problems we found regarding non-native English speakers. One usability study of an NHS website published in 2003 [17] found that major problems were often caused by specialized library terminology. This supports our findings regarding unfamiliar language and jargon. The few other usability studies of health-related web sites we uncovered dealt with online information for patients or the public.

Our results were used in discussions with The Cochrane Collaboration Steering Group and the publisher, Wiley-Blackwell, in order to develop and improve The Library web site. Other publishers of evidence-based content could use the more generic results to improve their own websites.

Searching (and finding): critical to evidence-based practice

The Cochrane Library site is not alone in having problems with findability. Results from usability tests of 217 web sites performed by Jakob Nielsen's team showed that search functionality and findability are the two largest categories of usability problems leading to task failure [7]. However, it is particularly ironic that a website built specifically to support evidence-based health care by synthesizing, organising and making accessible an overwhelming amount of health research should itself be perceived as overwhelming and difficult to navigate.

Discriminating design

In this study the non-native English speakers, though displaying no visible trouble reading English text, were at an extra disadvantage when trying to search. Their problems were related primarily to difficulty recalling and spelling query terms that resulted in relevant hits. Creating a reliable base of evidence is a task no organisation or country can solve alone – cross-national efforts are needed. Easy access to a body of high quality evidence should not be limited to native English-speaking participants. There is a wealth of technology that could be used to improve the user experience of searching for non-native English speakers. Spelling aids or query translation from other languages would be particularly helpful to these kinds of users. Automatic query expansion with synonyms (used by PubMed) could provide a better experience both for all searchers but would be particularly helpful for those with a limited English vocabulary.

Challenge – building a good mental model for evidence searching

Our findings revealed other challenges for designing good search functionality. In the Cochrane Database of Systematic Reviews, a precise query will result in only one or a few hits, as the underlying concept is one review per topic. However our participants' mental models of how search should function were based both on Google and PubMed, where simple queries produce a great number of results. The concept of a narrow search resulting only in a few hits is clearly still novel to many users and ways in which this can be made clearer need to be explored.

Challenge – building a good mental model of evidence-based information hierarchy

Our findings showed that systematic reviews can be confused with protocols and reports of clinical trials, even among experienced users who have a clear idea of the difference between these document types. This kind of misinterpretation may happen especially when different document types are mixed together in search results lists.

Different document types need to be distinguished from each other, both physically and visually – protocols should possibly be moved to a separate list. The importance of large clear labelling at the top of the individual documents enabling readers to easily distinguish between protocols, reviews and individual studies should also not be underestimated.

Appraising the source instead of the document

A related problem is the tendency for users to assume all Cochrane Library contents are Cochrane-approved. Most of our test persons seemed inclined to be satisfied with a quality assessment short-cut: making judgements about the trustworthiness of the *publishing source* rather than critically assessing individual documents of research as EBM teaching encourages. This inclination, when coupled with poor signposting on a site containing information of varying levels of editorial evaluation and research quality, leaves a gap wide-open for serious misunderstandings about the strength and quality of different pieces of evidence. Blind trust of a whole source is a complex labelling and branding problem and needs to be addressed by publishers on many levels.

Fragile credibility

Though Cochrane clearly enjoyed a high reputation among our participants, our study showed that even very small details can cause otherwise trusting users to suddenly question ownership and thereby credibility, such as an "About us" button leading to a page with a publishers' (unfamiliar) logo. While a large study from the Stanford Credibility project showed that consumers placed a lot of emphasis on the look of a site [18], a smaller parallel study showed that expert users tended to emphasize the reputation of the source when evaluating the trustworthiness of information found online [19]. Additionally it is important to follow the EBM principles of transparency and make it absolutely clear who is behind information that claims to be neutral and evidence-based.

This site is not for someone like me...

Many of our participants felt that *The Cochrane Library* site was for "*researchers*" or others with more knowledge than themselves, in part due to use of unfamiliar or academic jargon, but also connected to their failure to find relevant information. The feelings of ineptitude expressed by participants in this study is perhaps mirrored in the Australian study, where 41% of the participating physicians blamed their own limited search skills as impediments to making better use of research data. In fact, many of the problems our participants encountered were not due to their own lack of skills, but to design flaws that could be solved following usability heuristics [20] and research-based guidelines for web design [7,13,21] or implementing better search technology. It is also important to signal inclusiveness and relevance to other health care areas than just medicine. Clear signs of content produced for patient target groups could also serve to lower the perceived threshold for professionals.

Is valuable content enough?

Repeatedly we heard praise for the quality of content of this site. But frustration levels were very high, and several participants said they were ultimately too lazy to bother to use a site that made it so difficult for them. Information foraging theory describes user behaviour on the Internet as similar to wild animal's search for food: we want maximum benefit for a minimum of effort [22]. Jakob Nielsen points out that with the development of good search engines, it has become easier for information gatherers to move quickly between different hunting grounds, claiming that web sites should be designed less like big meals and more like tasty snacks, quick both to find and to eat [23]. A resource like Cochrane may be theoretically a great meal for a hungry animal, but too difficult to find and catch to be worth the effort, especially when less challenging prey is more easily available.

Limitations of this study

Our goal is to identify the emerging issues rather than to quantify them. In reporting results, we have therefore not emphasized frequencies of events. As our data set has not been designed to statistically represent a set of respondents, presenting numbers can be misleading [24].

The user tests were performed in a laboratory setting, and may not reflect actual behaviour or reactions from real-life situations. For instance, increased time pressure in clinical situations may result in even higher degrees of user frustration when an interface does not easily or intuitively produce quick results.

UK-based tests were held in the office of The Cochrane Collaboration, and this may have influenced the answers of participants regarding use and attitudes towards *The Cochrane Library* and Cochrane Reviews, despite our assurances that we were not connected with the design of the web site. Answers regarding familiarity and use of research were self-reported and not empirically validated.

The honeycomb model was not used to design the interview questions, only applied in retrospect to our data analysis. This may have affected the relevance of the data we collected to this model. On the other hand, this may have led to less "leading" questioning on our part.

The Cochrane Library, like most websites, is under continuous development/change, and several of the weaknesses we identified have since been improved.

Conclusion

Recommendations based on findings

Building web sites for evidence-based practice is not much different than building for good web usability in general. However, the consequences of not finding information or of finding the wrong information have potentially critical consequences. Health professionals' user experience of evidence-based online resources can be improved by applying the following principles:

- Follow existing usability heuristics and web usability guidelines, designing especially for findability through search engines, as well as for speed of use particularly important to health professionals.

- If resources are limited, focus on improving simple (non-advanced) search functionality, including technology that will help non-native English speakers.

- Drop "researcher" language and jargon to encourage use by health professionals.

- Don't assume users possess good mental models of evidence hierarchies. Make document types evident where possible – through information architecture, labelling, and search results design.

- Clearly mark the difference between quality-approved content and not quality-approved content.

- Ownership and authoring must be clear at all levels of the site for supporting and maintaining credibility.

Competing interests

CG is director of the Norwegian branch of the Nordic Cochrane Centre. JC is Co-ordinator & Managing Editor of The Cochrane Anaesthesia Review Group. CG and SR are involved in projects to improve summaries included in Cochrane Reviews. Test 2 was carried out in collaboration with Wiley-Blackwell, and partly funded by them.

Authors' contributions

SR conceived of and designed the study, carried out user testing, data analysis and interpretation, and drafting of the manuscript. CG helped design the study, carried out user testing, data analysis and interpretation, and drafting of the manuscript. JC recruited UK participants, carried out user testing, transcribed and coded the British tests, and commented on the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Appendix. Interview guide used in Test 2.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6947-8-34-S1.doc]

Acknowledgements

Irene Wiik Langengen recruited Norwegian participants and carried out user testing. Arild Bjørndal reviewed several versions of the manuscript.

References

1. Dickersin K, Straus SE, Bero LA: **Evidence based medicine: increasing, not dictating, choice.** *BMJ* 2007, **334(suppl_1):**s10.
2. Godlee F: **Milestones on the long road to knowledge.** *BMJ* 2007, **334(suppl_1):**s2-3.
3. Scott I, Heyworth R, Fairweather P: **The use of evidence-based medicine in the practice of consultant physicians. Results of a questionnaire survey.** *Aust N Z J Med* 2000, **30(3):**319-326.
4. Greenhalgh T: **How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses).** *BMJ* 1997, **315(7109):**672-675.
5. Collaboration TC: **The Cochrane Library.** [http://www.thecochranelibrary.com].
6. The Cochrane Collaboration: **Cochrane Handbook for Systematic Reviews of Interventions.** [http://www.cochrane-handbook.org].
7. Nielsen J, Loranger H: **Prioritizing Web Usability.** 1st edition. New Riders Press; 2006.
8. Kuniavsky M: **Observing the User Experience: A Practitioner's Guide to User Research.** In *Interactive Technologies* 1st edition. Edited by: Kaufmann M. Morgan Kaufmann; 2003:575.
9. Hassenzahl M, Tractinsky N: **User experience - a research agenda.** *Behaviour & Information Technology* 2006, **25(2):**91 -997.
10. Morville P: **User Experience Design.** :honeycomb model [http://www.semanticstudios.com/publications/semantics/000029.php].
11. **Morae** :Usability Testing for Software and Web Sites [http://www.techsmith.com/morae.asp].
12. **Web Accessibility Initiative** [http://www.w3.org/WAI]
13. Department of Health and Human Services: **Research-Based Web Design & Usability Guidelines.** Enlarged/Expanded edition. Washington, DC , U.S. Dept. of Health and Human Services; 2006.
14. Nielsen J, Tahir M: **Homepage Usability: 50 Websites Deconstructed.** New Riders Press; 2001:336.
15. Morville P: **Ambient Findability.** Sebastopol CA , O'Reilly Media, Inc.; 2005.
16. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS: **Evidence based medicine: what it is and what it isn't.** *BMJ* 1996, **312(7023):**71-72.
17. Ebenezer C: **Usability evaluation of an NHS library website.** In *Health Information and Libraries Journal Volume 20*. Health Libraries Group; 2003:134-142.
18. Fogg BJ, Cathy S, David RD, Leslie M, Julianne S, Ellen RT: **How do users evaluate the credibility of Web sites?** *Consumer Reports WebWatch* 2002 [http://www.consumerwebwatch.org/dynamic/web-credibility-reports-evaluate-abstract.cfm]. Stanford Persuasive Technology Lab, Cordura Hall 226, Stanford University, Stanford, Calif., 94305
19. Stanford J, Tauber E, Fogg BJ, Marable L: **Experts vs. online consumers: a comparative credibility study of health and finance web sites.** In *Consumer Reports WebWatch* Consumers Union of U.S. Inc.; 2002.
20. Nielsen J: **Ten usability heuristics.** [http://www.useit.com/papers/heuristic/heuristic_list.html].

21. Fogg BJ: **Stanford guidelines for web credibility. A research summary from the stanford persuasive technology lab.** [http://www.webcredibility.org/guidelines].
22. Pirolli P, Card S: **Information foraging.** *Psychological Review* 1999, **106(4):**643-675.
23. Nielsen J: **Information foraging: why google makes people leave your site faster.** [http://www.useit.com/alertbox/20030630.html].
24. Pope C, Ziebland S, Mays N: **Qualitative research in health care: analysing qualitative data.** *BMJ* 2000, **7227:**114-116.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6947/8/34/prepub

Rosenbaum, S. E. et al. (2010). User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. Journal of Clinical Epidemiology, 63 (6): 607-619.

Rosenbaum, S. E., C. Glenton & A. D. Oxman. (2010). Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *Journal of Clinical Epidemiology*, 63 (6): 620-626.

Evidence summaries tailored for health policymakers in low and middle-income countries

Sarah E Rosenbaum, Claire Glenton, Charles Shey Wiysonge, Edgardo Abalos, Luciano Mignini, Taryn Young, Fernando Althabe, Agustín Ciapponi, Sebastian Garcia Marti, Qingyue Meng, Jian Wang, Ana Maria De la Hoz Bradford, Suzanne Kiwanuka, Elizeus Rutebemberwa, Signe Flottorp, Andrew D Oxman

ABSTRACT

Objective

Systematic reviews can inform health policy decisions but are often not well suited to meet the information needs of policymakers. The SUPPORT collaboration developed a short summary format tailored to presenting results of relevant systematic reviews to policymakers in low and middle-income countries (LMIC).

Methods

We carried out 21 user tests in six countries examining users' experience of the summary format, made alterations based on these results, and checked our conclusions through 13 follow-up interviews. We supplemented these methods with advisory group feedback and working group workshops to generate feasible solutions to the problems uncovered by user testing.

Findings

Policymakers responded positively to a graded entry format (short summary with key messages upfront), but some struggled with comprehension of text and numbers. The three issues that were the most challenging in redesigning the evidence summaries were policymakers': 1) poor conceptual understanding of systematic reviews, 2) expecting information not found in systematic reviews, and 3) wanting shorter, clearer summaries. Solutions included adding 'information about the information' and formatting the text to make it easier to scan. In addition to the 'Key Messages', policymakers particularly valued the section on 'Relevance for LMIC', despite the lack of directly relevant evidence in the systematic reviews that were summarised.

Conclusion

Presenting evidence from systematic reviews to LMIC policymakers in clear concise summaries can render this information more understandable and useful, but careful attention needs to be paid to content framing and formatting in order to meet the information needs of this audience.

BACKGROUND

In order to maximize use of available resources, health policymakers in all countries need reliable up-to-date evidence about "what works"[1-3]. In low and middle-income countries (LMIC) the pressure to squeeze the most out of funds is even greater, as the gap between available resources and burden of preventable disease is much larger than in high-income countries (HIC).[4]

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 nonsystematic review.[5] 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.[3, 6, 7]

Many existing systematic reviews are relevant for the health care challenges of LMIC, reporting not only on the effects of clinical interventions but also on the effects of different arrangements for delivering, financing and governing health services. However these reviews are usually written for scientific or clinical audiences and are not necessarily well tailored for the information needs of policymakers.[8]

The Supporting Policy-relevant Reviews and Trials (SUPPORT) project was an international collaboration aiming to improve the use of reliable research evidence in LMIC policy and management decisions and to help fill in the gaps where there is a lack of reliable evidence. In this article we report on the development of SUPPORT summaries of systematic reviews for policymakers in LMIC. Our objective was to tailor a summary format that was sensitive to the needs of this particular audience.

METHODS

Selecting the reviews to be summarised

We searched The Cochrane Library, MEDLINE and EMBASE for relevant systematic reviews in September 2007 and repeated the search in 2008 and 2009. Selection criteria for SUPPORT summaries included:

- relevance to achieving maternal and child health goals in LMIC
- examination of at least one of the following:
    - the effects of interventions, services or programmes
    - the effects of delivery, financial or governance arrangements
    - the effects of strategies to implement change
- a methods section with explicit selection criteria

After a review was identified, two or more reviewers independently assessed its relevance based on the abstracts and, for those reviews that were potentially relevant, based on the full text.

Developing summary content
For this study, we selected five systematic reviews[9-13] covering different topics and including a variety of data types, such as dichotomous and continuous outcomes. We assessed the quality of these reviews using a checklist and extracted data about the topic and main results. The quality of evidence for the main comparisons in each review was graded using the GRADE system.[14, 15] Finally, we considered the applicability of the intervention to LMIC, possible impacts the intervention could have on equity, considerations related to scaling up of the intervention, and the need for further evaluation. These issues were chosen based on earlier research indicating their relevance to policymakers.[3, 16] The completed summary manuscripts were peer-reviewed by researchers and policymakers in LMIC, experts on the topic, and the authors of the reviews.

Developing the summary format
As a starting point for the summary format, we adopted the concept of a graded entry format [16, 17], where a one-page of key messages preceded an executive summary and a reference to the complete review. The executive summary deviated from a traditional academic format in the following ways:

- The "Methods" section was replaced by a short description of the characteristics of the underlying review.
- The "Discussion" section was replaced by a section called "Relevance for LMIC", addressing applicability of the evidence to LMIC, impacts on equity, considerations related to scaling up, and the need for further evaluation.

We used several methods to further develop the format and assess our progress (See flow chart, *Figure 1)*:

- *Working group workshops*, including three people with expertise in evidence dissemination (CG), information design (SR) and epidemiology (ADO) – to generate ideas and solutions to problems uncovered through feedback and testing
- *Advisory group feedback,* including multi-disciplinary LMIC researchers who would be authoring future summaries – to inform summary development from an authoring and production perspective
- *User testing* – to inform summary development from a LMIC policymaker perspective

A preliminary summary manuscript was authored by ADO and subsequently revised and reformated by the working group in a series of meetings.

Feedback from the advisory group was elicited at two project workshops, informing changes that resulted in the pilot summary version. Further feedback was gathered during a telephone conference following user testing and summary redesign.

Testing the summaries
The working group conducted three pilot user tests of the summaries with participants working in Norwegian government agencies involved in

development projects in LMIC. We made improvements on the summary based on results from these tests.

The advisory group then tested summaries with 18 policymakers in five different countries: Argentina(6), Colombia(3), Uganda(3), South Africa(3) and China(3). Spanish versions were used in Argentina and Colombia, while English versions were used in Uganda, South Africa and China. Participants included LMIC health policymakers and managers at different levels, purposively sampled by the advisory group. They were recruited by email and telephone, and were sent written information about the project by email or post prior to the interview.

The test method we employed was a think-aloud protocol, using a semi-structured interview guide. Participants were interviewed individually by a researcher in a session that lasted about one hour, with a second researcher observing and taking notes. Introductory questions covered the participants' education, employment, and experience using research, including familiarity with systematic reviews. The participants were presented with a list of the five summary topics from which they chose one to read at their own pace. Then they went through each section of the summary with the interviewer, who prompted them to think out loud about how they understood and experienced the presentation. The interview guide was based on a model for user experience (originally developed for interactive media) made up of seven facets: "findability", accessibility, credibility, usability, usefulness, desirability and value.[18] Usability was defined as easy to understand and quick to use. Since we were testing on paper, accessibility issues could not easily be explored and therefore were not included. Finally, the participants were asked for suggestions and additional comments.

The sessions were audio-taped and transcribed by the interviewers, who added their own notes. These transcriptions and notes were translated when necessary and sent to the working group, where they were merged into a single file. Audio tapes were subsequently erased, and participant identity was removed from the compiled results. Two researchers from the working group (SR and CG) performed separate analyses, identifying barriers or facilitators to favorable experiences of the summary according to the framework, as well as participants' suggestions for improvements. These analyses were then compared and reconciled by the two researchers. Barriers, facilitators and participant suggestions were sorted according to specific parts of the summary, such as the background section, or to general themes, such as language. They were also rated according to their severity. For instance, an observed barrier that hindered comprehension of the content was rated more severely than an explicit preference concerning color.

The working group used this analysis to develop a revised version of the summary format, altering content, presentation and format. We then presented the user test analysis and the new summary proposal to the advisory group in a telephone meeting and asked for responses and additional input. This resulted in minor changes that were made to create the final summary template.

After the summary was redesigned, we sent both the old and new summaries by post or email to the LMIC user test participants and briefly outlined the findings we had elicited from the user tests. We asked partipants to indicate which version of the summary they prefered and why, and also to give opinions on the accuracy of our test findings.

RESULTS

Results from the pilot tests were not contradictory with those from the final tests, so we have pooled all results below.

Only one test participant was employed full time in a school of medicine. The other 20 were primarily senior members of staff working in national or international health service or policy related work, such as health departments, health directorates, national insurance schemes, hospitals or aid organizations. Seventeen said they used research in their work, though elaboration on this question revealed that several seemed to define "research" very broadly to mean any information-gathering on a topic. Eighteen said they knew what a systematic review was, though six out of 21 were unfamiliar with Cochrane Reviews.

Of the six facets we explored from the user experience framework used to structure the interview, the most significant findings were in the categories of usefulness, usability (including issues regarding comprehension) and credibility.

Usefulness

Sixteen of the 21 participants reported that the summary would be useful for them if they were going to make a decision on this topic. The graded-entry format with key messages up front was perceived as being particularly useful:

> *"The title and the key messages are very useful. Short and concise."*

However many expressed that there was still a mismatch between the kind of content offered and their actual information needs:

> *"It explains that there is a high degree of satisfaction with what the nurse practitioners are doing compared to the doctors. But it doesn't say if they have limited the tasks that are performed… it doesn't say whether they are supposed to cover what the medical doctor or practitioner usually covers. And what sort of services? Is it general practice, is it in a hospital ward or what?"*

Some expectations about content seemed to stem from a poor understanding of what a systematic review was and what kind of information they could expect to find in such a summary. Unmet expectations included information outside the realm of a systematic review, such as recommendations, measurements of outcomes usually not included in a review, coverage of broader topics or more detailed information about local applicability and actual costs on a local level.

> *"From a manager's point of view, I would have liked to see information on cost."*

"(I) wouldn't ask this question. I prefer posing broader kinds of questions like 'convulsion treatment', but the question of the researcher of this review is for diazepam specifically…"

One participant suggested adding information about acceptability to different stakeholders.

Usability (ease of use and comprehension)

Despite the positive feedback on the front page, some still felt the summary (5-7 pages) was too long and too complex. They wanted a shorter, clearer presentation.

" Operational managers will be petrified. When I think summary, I think one page... I would not have time to read a long document even though I would want my work to be evidence based."

On the other hand, some felt the summary was not comprehensive enough.

"… this is not enough for a recommendation about mode of payment of physicians."

Although the authors had attempted to write in clear simple English (or Spanish), the presentation of findings in text and tables were still perceived by many as being too long and difficult. Eight participants found the tables difficult or confusing, and nine participants said the definitions of the concepts used in the table, such as GRADE or different presentations of risk, were not clear.

"This text (Summary of Findings) is quite difficult to interpret. This section would be very difficult to understand by people not trained in evidence-based medicine. Words like 'sample size', 'relative risk' would be difficult to interpret. These terms are too specific. It takes too much time to understand this section."

"…You know when you have a text with many figures and numbers it switches one off... "

Some tables ran over two pages, making them cumbersome to read. Abbreviations caused some confusion in both text and tables. In addition, we observed people comparing the summary of findings texts with the numbers in the tables, and becoming confused if these did not map up precisely with each other. For at least one participant (from China), reading information in the English language clearly posed a significant barrier. The term "scaling up" (from the Relevance section) was not always correctly understood as including economic considerations.

Credibility

Participants were asked early on in the interview if they would trust the credibility of the summary based on first impressions. Two participants

responded that they would because it seemed to be "well written". Twelve answered that they would trust it because they perceived it as coming from credible sources:

> *"I would trust in a report like this. It uses systematic reviews as source of information and I know that this kind of information is of high quality."*

> *"The references are clear as well as the source, that's the most important thing."*

However, not everybody understood that the summary stemmed from a systematic review, and some expressed confusion as to who was behind the summary as we had placed the partner logos on the last page. Also, some expressed reduced interest in the summary when they discovered that the evidence quality was low, when there was no evidence for important outcomes, or when the studies were old.

One participant who displayed familiarity with systematic reviews expressed confusion about how a good quality review could be compatible with low quality evidence for individual outcomes.

Value

Seventeen participants felt that a series of summaries like this one would be valuable to policymakers in a position like theirs, though one commented that they would need to be frequently updated to maintain their value.

Desirability

Fourteen participants said they liked the summary, particularly the front page with key messages and the 'Relevance for LMIC' section. Seven reacted positively to the 'Characteristics of the reviews' table.

> *"(I) like this chart; it makes clear what the review was looking for."*

Five specifically appreciated that the main title was framed as a question (Example: "Does pay-for-performance improve the quality of health care?").

> *"I liked the title as a question. It is motivating."*

"Findability"

When asked where they would expect to find this series of summaries, seven answered *"face-to-face meetings".* Many mentioned one or several centralized online sources, such as WHO, PAHO, Cochrane sites and health ministry or university web pages, though no one web site emerged as a primary preferred source.

> *"It is difficult to get reports like this, there are not many channels. So establishing better mechanisms for information dissemination is necessary."*

From test results to new summary design

A number of the findings pointed to obvious solutions, which we adopted:

- Continue to simplify all text and tables where possible
- Use consistent language when describing effect sizes and quality of evidence findings and quality in the text. (We chose to implement a structured set of terms and phrases developed in ongoing research aiming to improve Cochrane plain language summaries for consumers[19]. See *Table 1*.)
- Results in tables should correspond with results in the text
- Limit the number of tables in the summary
- No tables should run over two pages
- Eliminate all abbreviations
- Move partner logos to the front page
- Move date of summary publication to the front page
- Change heading text "Scaling up" to "Economic considerations"
- Add definitions of unfamiliar terms such as GRADE

However, three larger issues were more challenging to address: 1) poor conceptual understanding of systematic reviews, 2) expecting information not found in systematic reviews, and 3) wanting shorter, clearer summaries

The working group dealt with these problems in three ways. First, to address the problems of lack of conceptual understanding of systematic reviews and the wrong expectations about the kind of information this summary could provide, we added boxes of 'information about the information' (meta-information). Boxes on the front page included the following topics:

- "Who is this summary for?"
- "This summary includes:"
- "Not included in this summary:"
- "This summary is based on the following systematic review:"
- "What is a systematic review?"

On the following pages, we used similar boxes to add:

- "How this summary was produced" (increasing transparency)
- "Knowing what's not known is important" (addressing the issue of lack of evidence)
- "About quality of evidence (GRADE)"
- Information about the SUPPORT Collaboration, partners, links to a newsletter, glossary of terms, etc.

Next, to help meet the expectation of information not included in the review, we broadened the scope of the "Reference" section, adding references to information that helped to understand the problem, provided details about the interventions, or helped put the results of the review in a broader context. This section was renamed "Additional information".

The third change addressed the need for clearer and shorter text. Since the summary was already extremely condensed, it was difficult to make the text even shorter. Instead, we aimed to facilitate rapid scanning of the document, by

reformatting some of the text to make it easier to pick out the important parts more quickly. We made the following design changes:

- Separated '*findings'* in the text by reformatting them as bullet points and highlighted them with blue arrows to bring these parts more clearly to the foreground
- Divided the Relevance text into a table that separated '*findings'* from '*authors' interpretations'.*
- Moved the '*Characteristics of the review'* table to the background section. This made it possible to restrict the text in the background section to key information necessary to understand the objectives of the review.
- We also changed to a narrower font to prevent the summary from growing in length with the addition of content in boxes.

Additionally, in order to support summary authors in their efforts to create short, pertinent texts, we developed explicit instructions about what information to include and exclude from each section of the summary.

(Appendix 1: example of the SUPPORT Summary guideline template)

The advisory group agreed both on our interpretations of the user-test findings and the subsequent changes we made to the summary format.

Follow-up interviews

Thirteen of 21 participants responded to the follow-up questions comparing old and revised format. All clearly preferred the new format, explaining that they found it easier to read and more understandable. Reasons given for this were mainly the new front-page design and the addition of the 'information about the information' boxes:

> "*The content is presented in simple easy-to-understand language, especially the first page … The reference box on the right, on page one, is perfect as it tells you what to expect."*
>
> *"The information added in the boxes is very useful. I like very much i.e. 'what is included and what not'."*
>
> *"The table of results are much more understandable. It's great the description of GRADE. Now I can understand what it means."*

There was general agreement that our analysis of the problems was precise and that the new summary resolved the main issues. Two participants repeated earlier misgivings about missing content outside the scope of a systematic review. One participant felt that the tables were still confusing, as a definition of "relative risk" was still lacking.

(Figure 2: Final summary format, front page.)

DISCUSSION

User tests with health policymakers indicated that the graded entry format (one page of Key Messages followed by a short summary) is well suited to their needs. The sections “Key messages” and “Relevance for LMIC” were those parts of the summary participants showed most interest in. They had difficulty understanding risks presented in the tables, and were frustrated with text that seemed too long and complicated. Some revealed a poor understanding of what a systematic review was, and expected or wanted information not found in systematic reviews. There was also some confusion about the source of the summaries. We addressed all of these issues through alterations in the template’s content and design, in particular adding ‘information about the information’ and reformatting to increase ease of scanning the text. The advisory group and the participants who provided follow-up feedback supported this analysis and the subsequent changes to the summary.

Study strengths and weaknesses

The strength of our study is the participation of a wide range of policymakers from five LMIC and representing varying levels of decision making and familiarity with research evidence. In addition, we made use of a multi-disciplinary advisory group of LMIC researchers and summary authors. Although only two-thirds of the participants (13 of 21) responded to our follow-up interviews, they unanimously preferred the final redesigned version of the summary.

Several aspects of our study design may have weakened the study. The translation of transcripts from Spanish and Chinese to English introduced an extra filter between the interview text and the analysis, a factor that may have affected precise text interpretation. In addition, all of the user test interviewers were involved in summary production and participants knew this, potentially affecting their responses. Finally, summary topics were pre-selected and not necessarily matched to the current needs or interests of the policymakers who participated in the user testing. This may have affected both reading motivation and understanding of the material.

Other summaries and evaluations

A number of review-derived products for policymakers now exist, including summaries of systematic reviews, overviews of systematic reviews and policy briefs[20]. Several summaries of systematic reviews are targeted specifically at policymakers in LMIC. For example, Evidence Aid[21] provides two-page structured summaries of Cochrane Reviews for emergency settings. The South Africa Medical Journal (www.samj.org.za) publishes one-page summaries of studies and reviews for African settings.

Collections of summaries from high-income countries may also be relevant to LMIC policymakers. For instance the “Rx for Change” database[22] publishes summaries of systematic reviews of the effects of strategies to improve drug prescribing practice and drug use. "Evidence Boosts" [23] (The Canadian Health Services Research Foundation) summarizes healthcare issues where research indicates a preferred course of action in health services management and policy;

and The Policy Liaison Initiative[24] of the Australasian Cochrane Centre prepares web-based summaries for policymakers based on Cochrane Reviews.

One of the most important predictors of policymakers' use of systematic reviews is that they are easy to use.[25] However we uncovered few studies reporting evaluations of summary formats for policymakers. The studies we did find support our own findings. Lavis et al found that a graded-entry format and up-front take-home messages rendered HTA reports (Health Technology Assessments) more useful.[26] An evaluation of Evidence Aid summaries found that summaries might be more useful in catastrophe situations if they were based on broader topics rather than single reviews and that language should be tailored to non-clinical audiences.[27] Both studies found that content which helped users contextualize the evidence (discussion of applicability and relevance) was particularly valuable.

Shorter messages or text that is quicker to scan?

One of the overriding findings was the desire for clear and short messages. This has been found in other studies where policymakers have been interviewed for their preferences regarding presentation of research.[16, 28, 29] 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, other devices than text editing must be used, such as graded entry structuring of the text with a first page summary of key messages. In recent years, research on use of web sites has taught us much about how people scan, as opposed to how they read.[30, 31] This kind of knowledge can be brought into contexts where readers have little time for long text. Bullet lists, shorter paragraphs and judicious use of headings are known devices for improving the ease of scanning text.[30]

Supporting better comprehension

We uncovered a number of problems due to poor comprehension of numbers and statistics. Studies show that even highly educated people struggle to understand risks.[32] Statistical literacy is "an understanding of concepts such as chance and uncertainty, sampling variability, margins of error, and randomization in clinical trials, and the ability to use such concepts to evaluate scientific information."[33] Appreciation of the value (and limitations) of systematically reviewed evidence is dependant on a basic understanding of these 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 the main messages. However correct comprehension is not only dependent on the skills and knowledge of the reader, but also on the characteristics of the information [33]. 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 "About quality of evidence") may help replace frustration with reflection, for instance in the case of weak or missing evidence.

Problem structuring rather than problem solving

Systematic reviews attempt to answer narrowly defined scientific questions, for instance whether or not an intervention has a specific effect. Health care policy issues are larger problems that need answers beyond "Will it work?" These include answers to questions such as "Will it work here?" "What are the consequences?" "What will it cost?" A single summary published for LMIC generally cannot provide answers to these other kinds of questions for a multitude of specific settings. However, 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 was done in the section on 'Relevance for LMIC', where findings and interpretations related to applicability, equity and cost were outlined. Despite lack of specific answers in the text, this information was found very useful by policymakers.

This indicates that there may be value in supporting problem structuring even when it is not possible to provide specific solutions. Earlier research supports this finding, suggesting that though research findings may not be of direct instrumental use in policymaking, they may be of conceptual use.[34-37] When the evidence quality is too weak to provide conclusive 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 policymaking processes.

CONCLUSIONS

Systematic reviews are an important resource, but policymakers are not familiar with them and they are not easily accessible. Summaries of systematic reviews can help address these problems. In order to be useful to LMIC policymakers summaries must be perceived as being clear and easy to read or scan quickly. They can also be designed to help readers better understand the nature of information from a systematic review and its applicability to policy decisions in LMIC. This may help to improve the usefulness of evidence from systematic reviews for health policy decisions.

ACKNOWLEDGEMENTS

The SUPPORT Collaboration was funded by the European Commission's 6th Framework Programme Priority FP6-2004-INCO-DEV-3.

SUPPORT summaries can be downloaded free at www.support-collaboration.org.

REFERENCES

1. Task Force on Health Systems Research. Informed choices for attaining the Millennium Development Goals: towards an international cooperative agenda for health-systems research. 2004;364:997-1003.
2. Travis P, Bennett S, Haines A, Pang T, Bhutta Z, Hyder A, et al. Overcoming health-systems constraints to achieve the Millennium Development Goals. The Lancet. 2004;364:900-06.
3. Lavis J, Posada F, Haines A, Osei E. Use of research to inform public policymaking. The Lancet. 2004 Oct 30-Nov 5, 2004;364(9445):1615.
4. The Millenium Development Goals Report 2008. New York: United Nations2008.
5. Greenhalgh T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). BMJ. 1997 September 13, 1997;315(7109):672-5.
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(2).
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. Waters H, Hatt L, Peters D. Working with the private sector for child health. Health Policy Plan. 2003 Jun;18(2):127-37.
10. Briggs CJ, Garner P. Strategies for integrating primary health services in middle- and low-income countries at the point of delivery. Cochrane Database Syst Rev. 2006;19(2):CD003318.
11. Gosden T, Forland F, Kristiansen IS, Sutton M, Leese B, Giuffrida A, et al. Impact of payment method on behaviour of primary care physicians: a systematic review. J Health Serv Res Policy. 2001 Jan;6(1):44-55.
12. Beney J, Bero LA, Bond C. Expanding the roles of outpatient pharmacists: effects on health services utilisation, costs, and patient outcomes. Cochrane Database Syst Rev. 2000;3(3):CD000336.
13. Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, et al. Lay health workers in primary and community health care. Cochrane Database Syst Rev. 2005;25(1):CD004015.
14. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004 June 19, 2004;328(7454):1490-.
15. GRADEpro - Information Management System (IMS). [07 January 2009]; GRADEpro (GRADEprofiler) is the software used to create Summary of Findings (SoF) tables in Cochrane systematic reviews. Sofware download and technical information, support, feedback and other resources.]. Available from: http://www.cc-ims.net/gradepro.
16. Lavis J, Davies H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. J Health Serv Res Policy. 2005 Jul;10 Suppl 1:35-48.
17. Canadian Health Services Research Foundation. Reader-Friendly Writing - 1:3:25: Canadian Health Services Research Foundation,2001.

18. Morville P. User Experience Design. [website]: Sematic Studios LLC; 2004 [updated June 21, 2004April 09, 2008]; honeycomb model]. Available from: http://www.semanticstudios.com/publications/semantics/000029.php.
19. Glenton C, Santesso N, Rosenbaum SE, Nilsen E, Rader T, Ciapponi A, et al. Presenting the results of systematic reviews to patients and
the public in four countries: A qualitative study. In press. [research]. 2009.
20. Lavis J. Supporting the Use of Systematic Reviews in Policymaking. PLoS Med. 2009;in press.
21. Evidence Aid Summaries. The Cochrane Collaboration; [cited 2009 August 29]; Available from: www.cochrane.org/evidenceaid/index.htm.
22. Rx for Change [database on the Internet]. Canadian Agency for Drugs and Technologies in Health. [cited August 29, 2009]. Available from: www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions/.
23. Evidence Boost. Canadian Health Services Research Foundation; [cited 2009 August 29]; Available from: www.chsrf.ca/mythbusters/eb_e.php.
24. Policy Liaison Initiative. Australasian Cochrane Centre; [cited 2009 August 29]; Available from: www.cochrane.org.au/ebpnetwork/.
25. Dobbins M, Cockerill R, Barnsley J. Factors affecting the utilization of systematic reviews. International Journal of Technology Assessment in Health Care. [research article]. 2001;17(2):203-14.
26. Lavis JN, Wilson MG, Grimshaw J, Haynes RB, Ouimet M, Rainan P, et al. Towards Optimally Packaged and Relevance Assessed Health Technology Assessments. research study2007 13 May, 2007.
27. 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.
28. 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(3):156-63.
29. 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.
30. Morkes J, Nielsen J. Concise, SCANNABLE, and Objective: How to Write for the Web. 1997 [cited 2009 April 25]; Web page article reporting research results]. Available from: http://www.useit.com/papers/webwriting/writing.html.
31. Aula A, editor. Enhancing the Readability of Search Result Summaries. HCI 2004: Design for Life; 2004.
32. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001 Jan-Feb;21(1):37-44.
33. Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc. 2007 Nov-Dec;14(6):713-21.
34. Innvær S, Vist GE, Trommald M, Oxman AD. Health policy-makers' perceptions of their use of evidence: a systematic review. Journal of Health Services Research & Policy. [systematic review]. 2002 October 2002;7(4):239-44.

35. Lomas J. Improving Research Dissemination and Uptake in the Health Sector: Beyond the Sound of One Hand Clapping. Policy Commentary: Mc Master University Centre for Health Economics and Policy Analysis 1997 Novermber 1997.
36. Weiss CH. The Many Meanings of Research Utilization. Public Administration Review. 1979 Sep.-Oct., 1979;39(5):426-31.
37. Nutley SM, Walter I, Davey HTO. Using Evidence. How research can inform public services. Bristol: The Policy Press; 2007.

Figure 1: Flow chart

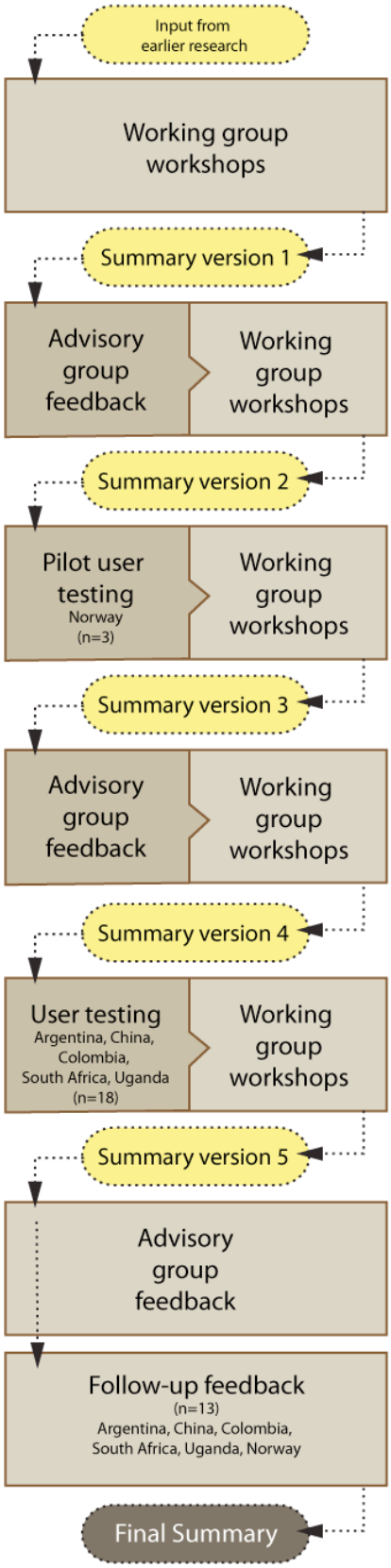
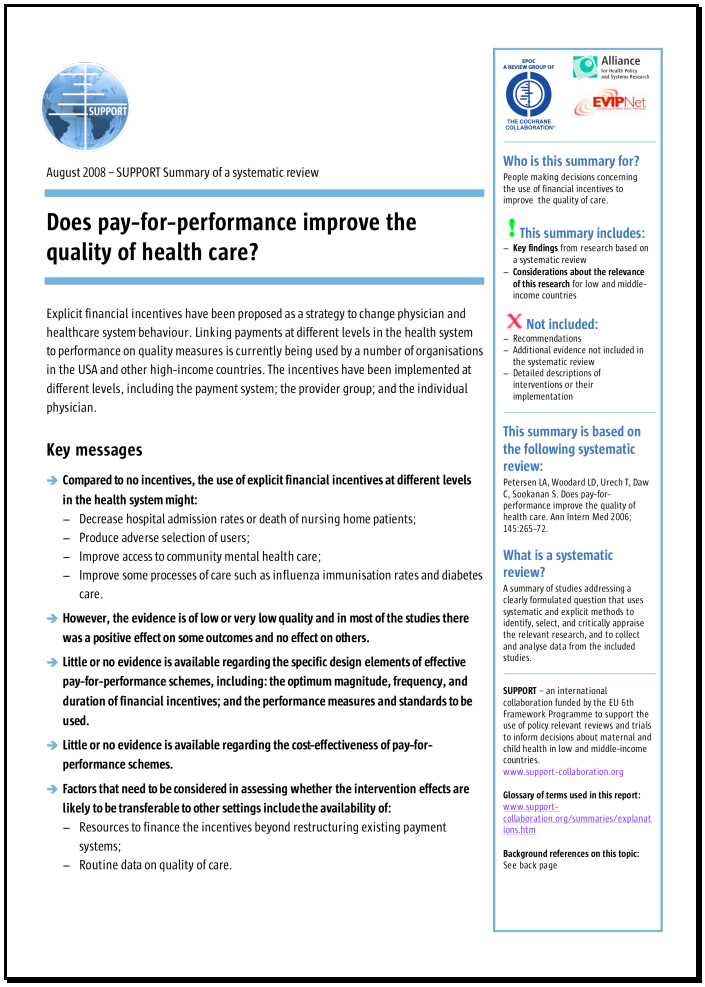


Figure 2: Example of final summary, front page.



August 2008 – SUPPORT Summary of a systematic review

Does pay-for-performance improve the quality of health care?

Explicit financial incentives have been proposed as a strategy to change physician and healthcare system behaviour. Linking payments at different levels in the health system to performance on quality measures is currently being used by a number of organisations in the USA and other high-income countries. The incentives have been implemented at different levels, including the payment system; the provider group; and the individual physician.

Key messages

- **Compared to no incentives, the use of explicit financial incentives at different levels in the health system might:**
  - Decrease hospital admission rates or death of nursing home patients;
  - Produce adverse selection of users;
  - Improve access to community mental health care;
  - Improve some processes of care such as influenza immunisation rates and diabetes care.
- **However, the evidence is of low or very low quality and in most of the studies there was a positive effect on some outcomes and no effect on others.**
- **Little or no evidence is available regarding the specific design elements of effective pay-for-performance schemes, including: the optimum magnitude, frequency, and duration of financial incentives; and the performance measures and standards to be used.**
- **Little or no evidence is available regarding the cost-effectiveness of pay-for-performance schemes.**
- **Factors that need to be considered in assessing whether the intervention effects are likely to be transferable to other settings include the availability of:**
  - Resources to finance the incentives beyond restructuring existing payment systems;
  - Routine data on quality of care.

Who is this summary for?
People making decisions concerning the use of financial incentives to improve the quality of care.

This summary includes:
- **Key findings** from research based on a systematic review
- **Considerations about the relevance of this research** for low and middle-income countries

Not included:
- Recommendations
- Additional evidence not included in the systematic review
- Detailed descriptions of interventions or their implementation

This summary is based on the following systematic review:
Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care. Ann Intern Med 2006; 145:265-72.

What is a systematic review?
A summary of studies addressing a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise the relevant research, and to collect and analyse data from the included studies.

SUPPORT – an international collaboration funded by the EU 6th Framework Programme to support the use of policy relevant reviews and trials to inform decisions about maternal and child health in low and middle-income countries.
www.support-collaboration.org

Glossary of terms used in this report:
www.support-collaboration.org/summaries/explanations.htm

Background references on this topic:
See back page

Table 1: Standardized language for describing effect sizes and quality of evidence[19]

	Important benefit/harm	***Less important benefit/harm***	***No important benefit/harm or null effect***
High quality evidence	"Increases/decreases"	"Increases/decreases slightly"	"Makes little or no difference"
Moderate quality evidence	"Probably increases/decreases"	"Probably increases/decreases slightly"	"Probably makes little or no difference"
Low quality evidence	"May increase/decrease"	"May increase/decrease slightly"	"May make little or no difference"
Very low quality evidence	"It is not known/we are uncertain whether [intervention] increase/decrease [outcome]"		

Appendix 1: Example of author's template.

Norsk samfunnsvitenskapelig datatjeneste AS
NORWEGIAN SOCIAL SCIENCE DATA SERVICES

Hans Holmboes gate 22
N-5007 Bergen
Norway
Tel: +47/ 55 58 21 17
Fax: +47/ 55 58 96 50
nsd@nsd.uib.no
www.nsd.uib.no
Org.nr. 985 321 884

ARKIV 427
Dok nr 3

Sarah Rosenbaum
Nasjonalt kunnskapssenter for helsetjenesten
Postboks 7004 St. Olavs plass
0130 OSLO

Vår dato: 19.08.2005 Vår ref: 200501233 LT /RH Deres dato: Deres ref:

KVITTERING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 15.07.2005. Meldingen gjelder prosjekt:

13088	*Brukertest Cochrane Library*
Behandlingsansvarlig	*Nasjonalt kunnskapssenter for helsetjenesten, ved institusjonens øverste leder*
Daglig ansvarlig	*Sarah Rosenbaum*

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres slik det er beskrevet i vedlagte prosjektvurdering. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, http://www.nsd.uib.no/personvern/register/

Personvernombudet vil ved prosjektets avslutning, 01.01.2011, rette en henvendelse angående status for behandling av personopplysninger..

Vennlig hilsen

Bjørn Henrichsen

Lis Tenold

Kontaktperson: Lis Tenold tlf: 55 58 33 77

Prosjektvurdering

Daglig ansvarlig

Sarah Rosenbaum
Nasjonalt kunnskapssenter for helsetjenesten

Postboks 7004 St. Olavs plass
0130 OSLO

13088 **Brukertest Cochrane Library**

FORMÅL
Formålet med prosjektet er å gjennomføre en brukertest av enkelte nettsteder som en del av et større arbeid med å se på hvordan helsearbeidere bruker ulike informasjonsressurser. Formålet med brukertesten er å se hvordan ansatte i helsetjenesten med liten eller ingen kjennskap til Cochrane Library finner, bruker og opplever dette nettstedet. Undersøkelsen vil antyde om brukergrensesnittet er tilstrekkelig tilpasset denne målgruppen, og identifisere områder for potensiell forbedring av brukeropplevelsen.

UTVALG - REKRUTTERING - FØRSTEGANGSKONTAKT
Utvalget omfatter 15 personer som er ansatt i helsetjenesten eller forvaltningen, og som har lett etter helserelatert informasjon på Internett. Rekrutteringen skjer ved at prosjektleder forsøker å rekruttere deltagere som tidligere har vært på kunnskapshåndteringskurs hos dem, fordi en vet disse vil ha en viss grunnkunnskap om innholdet på nettstedet. Hvis for mange melder seg, velger en en blanding av menn og kvinner som representerer ulike yrkesgrupper.

Førstegangskontakt med potensielle deltagere skjer på to forskjellige måter:
1) Generell forespørsel på 3 diskusjonslister (e-post) hvor prosjektleder ber interesserte om å melde tilbake til Nasjonalt kunnskapssenter for helsetjenesten.
2) Direkte forespørsel til noen ansatte i Sosial- og helsedirektoratet som har deltatt på kunnskapshåndteringskurs.

INFORMASJON OG SAMTYKKE
Det gis skriftlig informasjon på e-post og innhentes skriftlig samtykke for deltakelse.

METODE FOR DATAINNSAMLING
Opplysningene samles inn gjennom tester. Testleder og testpersonen sitter sammen på et rom hvor det finnes en datamaskin utstyrt med brukertest-programvare og web-kamera + mikrofon. Brukertest-programmet fanger alle bevegelser fra skjermen på et videospor, og syr dette synkront sammen med videoopptak fra web-kamera og lydspor fra mikrofonen. Testlederen innleder med et par spørsmål, deretter blir testpersonen bedt om å gjennomføre enkelte oppgaver på det aktuelle nettstedet. Testpersonene blir oppmuntret til å snakke høyt om hva de tenker mens de gjennomfører oppgavene. På denne måten fanges både billedopptak av skjermen mens testpersonen gjennomfører oppgavene, ansiktsuttrykket, og stemmelyden til testpersonen, som til sammen danner et godt inntrykk av hvordan nettstedet blir opplevd av vedkommende under bruk. Etter testen: Testleder avslutter med et par spørsmål.

Prosjektleder opplyser at testmanus ikke er ferdig utarbeidet enda, men vil inneholde følgende:
Før-test prat:
Innledende spørsmål:
Hva er ditt yrke? Bruker du Internett for å finne helserelatert kunnskap i forbindelse med arbeidet ditt? Kjenner du til the Cochrane Library? Har du besøkt dette nettstedet før? Hvor ofte? Når besøkte du det sist? Hva gjør du der/hva pleier du å bruke the Cochrane Library til? Er det andre lignende nettsteder eller databaser du besøker/bruker ofte i forbindelse med jobben?

Bruk av nettstedet:
Testoppgaver vil være skreddersydde slik at de er mest mulig relevante for den enkelte testpersonen. Dette er viktig å få til for at denne laboratorie-testen skal ligne mest på en virkelig situasjon hvor vedkommende selv er motivert til å finne ut av noe. Innhold i oppgavene vil formes på bakgrunn av telefonsamtalen tidligere med den enkelte. Men alle oppgavene vil fokusere på felles problemstillinger som har med søk, logg inn og evaluering av søkeresultatene å gjøre, i tillegg til en generell oppfatning av nettstedets omfang og innhold.

Avsluttende spørsmål:
Av typen: Hvordan synes du dette nettstedet var å bruke? Synes du det var enkelt eller vanskelig å finne frem? Likte du dette nettstedet? Hva likte du/likte du ikke spesielt? Var det noe du savnet? Vil du anbefale kollegaer å bruke dette nettstedet?

REGISTRERING OG OPPBEVARING
Materialet blir redigert på en pc frakoblet nettverket, og som er både beskyttet med brukernavn og passord samtidig som at det står i et låst rom. Opptakene blir bevart på en lokal ekstra harddisk under redigering, ikke på en felles server. Resultatene blir til slutt lagret på en DVD og slettet fra maskinen. Navn og personlig kontaktinformasjon blir oppbevart på en annen pc og aldri sammen med opptakene eller knyttet til opptakene på annen måte.

Opptakene oppbevares på en separat pc fra navn og kontaktinformasjon. Kontaktinformasjon tas vare på for å ha evt. mulighet til å be om utvidet samtykke for bruk.

Innsamlede opplysninger anonymiseres ved prosjektslutt, senest 01.01.2011. Med anonymisering innebærer at navnelister slettes/makuleres, og evt. kategorisere eller slette indirekte personidentifiserbare opplysninger samt at opptak slettes. Prosjektet vil ikke gjøre bruk av elektronisk sammenstilling av personregistre.

RISIKOVURDERING
Vurderinger om gjenkjenning:
Dersom sikkerhetstiltakene brytes eller noen av testpersonene på annen måte blir gjenkjent, vil ikke det føre til tap av integritet eller anseelse av testpersonen. Hovedmålet er å teste nettstedet, ikke brukere. Både spørsmålene og presentasjonen av resultatene vil legge vekt på problemer i brukergrensesnittet som fremstår som barrierer for den enkelte, ikke på testpersonens ferdigheter.

Hvorfor det er nødvendig med opptak:
Det er fullt mulig å utføre slike tester uten å ta opp lyd og bilde, men i stedet bruke et observasjonsrom. Denne teknikken er veldig effektfull når de som har ansvar for et nettsted (ledelse, teknologer, designere, redaktører) kan møte opp og sitte sammen i observasjonsrommet mens testen pågår. Men i dette tilfellet tilhører nettstedet en internasjonal organisasjon, og innholdet genereres av forskere i mange land. "Ledelsen" er et stort antall personer som befinner seg spredt over hele verden. Nettstedet publiseres av en tredje part, et forlagshus som også er å finne på flere kontinenter. Det er helt umulig å samle noen interessenter under et tak for å observere en brukertest sammen, derfor brukes videoopptak som kan vises når mange interessenter er samlet på konferanse.

KOMMENTAR
Personvernombudet for forskning finner opplegget for gjennomføringen av prosjektet tilfredsstillende. Videre finner vi informasjonsskrivene godt utformet. Det gis informasjon om alle deler av prosjektet. Videre innhentes det eksplisitt samtykke for at prosjektleder kan fremvise en redigert utgave av opptakene på en konferanse i oktober 2005.

Datamaterialet oppbevares ved arkivet til Nasjonalt kunnskapssenteret for helsetjenesten. Da testen kan muligens bli en pilot for et doktorgradsprosjekt, vil datamaterialet oppbevares lenge nok til at det kan bli brukt som en del av doktorgraden. Disketten og alle personopplysningen slettes etter 5 år. Det er gitt informasjon og innhentet skriftlig samtykke for dette.

Co-authorship: description of roles

Article 1: *User experiences of evidence-based online resources for health professionals: User testing of The Cochrane Library.*

Authors (in correct order)	**Role**
Sarah Rosenbaum	Corresponding author Conceived of and designed the study; carried out user testing (collection, transcription, analysis and interpretation of the data); drafted the manuscript.
Claire Glenton	Helped design the study; carried out user testing (collection, transcription, analysis and interpretation of the data); helped draft the manuscript.
Jane Cracknell	Recruited UK participants; carried out user testing (data collection, transcribed and coded the British tests); commented on the manuscript.
All	All authors read and approved the final manuscript.

Article 2: *User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews.*

Authors	**Role**
Sarah Rosenbaum	Corresponding author Designed the user test part of the study, carried out user testing (collection, transcription, analysis and interpretation of the data); interpreted stakeholder feedback; participated in working group workshops; designed tables; drafted the manuscript.
Claire Glenton	Helped design the user test part of the study, carried out user testing (collection, transcription, analysis and interpretation of the data); interpreted stakeholder feedback; participated in working group workshops; helped draft the manuscript.
Hilde Kari Nylund	Helped carry out user testing (collection, transcription, analysis and interpretation of the data); participated in working group workshops; commented on the manuscript.

Andrew D. Oxman	Conceived the study; designed the stakeholder feedback part of the study; collected and interpreted stakeholder feedback; participated in working group workshops; commented on the manuscript.
All	All authors read and approved the final manuscript.

Article 3: Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information.

Authors	**Role**
Sarah Rosenbaum	Corresponding author Helped with study design and data collection. Interpreted data. Drafted the manuscript.
Claire Glenton	Helped with study design, data collection and data interpretation. Helped draft the manuscript.
Andrew D. Oxman	Conceived the study; designed the data collection; collected and interpreted data; commented on the manuscript.
All	All authors read and approved the final manuscript.
Non-authors	Jan Ødegaard-Jensen carried out the statistical analysis. Arild Bjørndal proofread the manuscript.

Article 4: Evidence summaries tailored for health policymakers in low and middle-income countries.

Authors	**Role**
Sarah Rosenbaum	Corresponding author Designed the study; carried out user testing pilot (collection, transcription, analysis and interpretation of the data); analyzed and interpreted data from the user tests; collected, analyzed and interpreted stakeholder feedback; participated in working group workshops; designed summaries; drafted the manuscript.
Claire Glenton	Helped design the study, carried out user testing pilot (collection, transcription, analysis and interpretation of the data); participated in working group workshops; helped draft

	the manuscript.
Charles Shey Wiysonge	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Edgardo Abalos	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Luciano Mignini	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Taryn Young	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Fernando Althabe	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Augustín Ciapponi	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Sebastian Garcia Marti	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Qingyue Meng	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Jian Wang	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Ana Maria De la Hoz Bradford	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Suzanne Kiwanuka	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Elizeus Rutebemberwa	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Signe Flottorp	Carried out user tests (collection and transcription of the data), commented on the manuscript.
Andrew D Oxman	Conceived the study; participated in working group workshops; commented on the manuscript.
All	All authors read and approved the final manuscript.

Support summary user testing

Test person: 0
Soundtrack nr. 00x
Place: city, country
Date: month, year
Researcher: your name, institution

Check list, for facilitator:
- 4 summaries
- title sheet for choosing one summary
- tape recorder w/extra batteries

for observer/note taker:
- copy of 4 summaries
- note taking sheet

Introduction

Go through information sheet.

What we are testing and why
We are going to look at the usability of some material that is under development for the SUPPORT project. SUPPORT is an international collaboration funded by the EU 6th Framework Programme to support policy relevant reviews and trials to inform decisions about maternal and child health in low and middle-income countries. You are one of several people that we are collecting feedback from in this form around the world. We'll use this feedback to improve the material, so that it will be as good as possible and easy to use for people with backgrounds similar to yours.

What will we be doing?
First we ask you some background questions. Then you will be shown some material and I'll be asking you questions about it. We want you to answer from your own perspective, not on the behalf of other people.

The session will be taken up on tape if that's ok with you. The tapes will only be used for transcribing, and will be erased afterwards. The whole process will take about one hour.

About user testing
From our experience, we are fairly certain that things you find difficult to understand, other people will also find difficult. So we can use this information to make the material better. We are out to find what works well and what works less well, both regarding content, use of language or terminology, as well as presentation and formating. We very much want to hear YOUR OWN OPINION, so there is no right or wrong answer to anything we ask. We are not testing you, we are testing our material.

Who is doing what
I will be leading the test, xxxx will be taking notes.

- Do you have any questions about the project?

Turn on recorder.

Background questions

A	What is your educational background and your current position?	

B	Do you sometimes read research results in connection with your work?	☐ yes ☐ no

C	Think of an example of a recent policy decision that you recently were involved in - what sort of information did you use, and where did you look for information? (websites, journals, colleagues, etc)

D	Do you know what a systematic review is?	☐ yes ☐ no

E	Have you read a systematic review or part of one?	☐ yes ☐ no

F	Have you heard of Cochrane Reviews?	☐ yes ☐ no
G	How familiar are you with them?	
	☐ not familiar ☐ read/browse seldom ☐ read/browse now and then ☐ read/browse regularly ☐ author or co-author of a Cochrane Review	

Questions about the summary

A short bit of repetition before we begin.

No right or wrong answer

You are not being tested, it is our material we are testing. There are no right or wrong answers to our questions. If you think something is easy or difficult, clear or confusing, if you understand or don't understand, we just want to know about it.

Think out loud

Think out loud. Tell me what you are thinking, what you see, what you find confusing or surprising, even the least little bit. For instance:

- What you are looking at, describe your experience of it.
- If you are unsure about anything
- If you are surprised by anything
- If there are things you don't understand, just say "I don't know what this means..."

My role

My role is to ask questions. But, since it is your opinion we are interested in, I will be otherwise saying as little as possible. You can ask me questions, but I won't be answering them. If you like, I can answer them as well as I can when we are finished.

First impressions

spontaneous first impression

1	**Before showing the report:** **I'm going to show you a report that is an example of a series that is being developed. I want you to imagine that you found it as a link on a website that you often visit, and that you chose to download it.** **Hand them the sheet with report titles.** **Here are the titles of four different topics. Please choose the title that interests you most right now.** **Before I give you the report, I want your first immediate impression, your spontaneous reaction to it when I show you this. Don't think, just tell me the first thing that comes into your head when you see it.** **Give them the report.** **What is your first spontaneous reaction?**

Overview, quick understanding of the structure

2	**Without reading in too much detail (we'll go much more into depth in a minute), do you get any idea of what information you might find in this report by glancing at it for a moment or two?**

How would they normally read a report?

3	Show me how you would normally go about reading a report like this. Where would you start, what would you look for first etc? How long would you normally use? (Published on web site, pdf, print out, read on screen).

Credible

4	You've just had a brief look at this report. Based on this, could you say anything about your impression of the credibility of this report? Do you think you would trust this information? Why, why not?

Now ask them to read the report, using as much or little time as they like. Remind them that you will not be asking exam-like questions afterwards. Go into another room while they do this on their own.

Usable

5	Now I'd like you to go through each part of the report, every element, and describe what your understanding of it is. Start up here at the top of the first page and go through each part of the whole publication, and just tell me if things are clear to you or unclear, or if there is anything missing you might be looking for....

5a	First page with title, logos, key messages
5b	Background
5c	Summaries: text
5d	Summaries: text with "Summary of Findings" table looking to find out if we need adjusted versions of the table for this group - walk through each element of table - ask them to explain how they interpret what is presented, repeat in their own words how they understand (or don't understand) the results - go through 1 continuous, 1 dichotomous outcome, 1 empty row if relevant
5e	Comments on relevance
5f	Characteristics of the review

5g	**Back page**

If there is time: Show more complex Summary of Findings table
If they have chosen a report with a simple table, show them a table that is more complex from a different report, and go through it in a similar fashion, row by row. We are particularly interested in their first impressions of the table, whether they would bother to read it, and in detail to learn how they understand the results in columns 2 and 3.

If not time, continue below:

Understandable (self-experienced)

6	**Do you think this report was generally easy or generally difficult to understand? Explain...**

Useful

7	**Would this report would be useful for you if you were going to make a decision about health care policy on this topic?**

Desirable

8	**To the degree you can "like" a report, did you like this report or not like it? Explain…** **If you could change it in any way (content, language, or formatting) what would you change?**

Valuable

9	**Do you think a series of these types of reports would be valuable for people in positions similar to yours?**

Suggestions for increasing value

9	**Could they be made more valuable for you?** **If it was up to you to make changes, what would you change?** **(Content, language, formatting)?**

Findable

10	**Where would you expect to find reports like this? (or if you had heard that a series of reports like this existed, how would you go about finding them?)** **Do you have any specific suggestions for spreading them to relevant audience?**

That was all the questions I have about the report, but before we finish I'd just like to ask about the test itself:

Improving our test?

Do you have any suggestions as to how we might have done this test better, for instance the information you received, etc.?

Thank you, that was all, we are finished.

Summary of findings:

Compression stockings compared with no compression stockings for people taking long flights

Patients or population: Anyone taking a long flight (lasting more than 6 hours)
Settings: International air travel
Intervention: Compression stockings[1]
Comparison: Without stockings

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Without stockings	**With stockings**				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies.
Symptom-less deep vein thrombosis	**Low risk population** [2]		**RR 0.10** (0.04 to 0.25)	2637 (9 studies)	⊕⊕⊕⊕ High	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population [2]					
	30 per 1000	**3 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**RR 0.45** (0.18 to 1.13)	1804 (8 studies)	⊕⊕⊕◯ Moderate[3]	
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6 to 9**.	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI −4.5 to −4.9).		1246 (6 studies)	⊕⊕◯◯ Low[4]	
Pulmonary embolus	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed pulmonary embolus in these studies.
Death	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants died in these studies.
Adverse effects	See comment	See comment	Not estimable	1182 (4 studies)	See comment	The tolerability of the stockings was described as very good with no complaints of side effects in 4 studies. [5]

*The basis for the **assumed risk** is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the intervention group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see explanations)

[1] All the stockings in the 9 trials included in this review were below-knee compression stockings. In four trials the compression strength was 20-30 mm Hg at the ankle. It was 10-20 mm Hg in the other four trials. Stockings come in different sizes. If a stocking is too tight around the knee it can prevent essential venous return causing the blood to pool around the knee. Compression stockings should be fitted properly. A stocking that is too tight could cut into the skin on a long flight and potentially cause ulceration and increased risk of DVT. Some stockings can be slightly thicker than normal leg covering and can be potentially restrictive with tight foot wear. It is a good idea to wear stockings around the house prior to travel to ensure a good, comfortable fitting. Stockings were put on 2 to 3 hours before the flight in most of the trials. The availability and cost of stockings can vary.

[2] Two trials recruited high risk participants defined as those with previous episodes of DVT, coagulation disorders, severe obesity, limited mobility due to bone or joint problems, neoplastic disease within the previous two years, large varicose veins or, in one of the studies, participants taller than 190 cm and heavier than 90 kg. The incidence for 7 trials that excluded high risk participants was 1.45% and the incidence for the 2 trials that recruited high-risk participants (with at least one risk factor) was 2.43%. We have rounded these off to 10 and 30 per 1,000 respectively.

[3] The confidence interval crosses no difference and does not rule out a small increase.

[4] The measurement of oedema was not validated or blinded to the intervention. All of these studies were conducted by the same investigators.

[5] None of the other trials reported adverse effects, apart from 4 cases of superficial vein thrombosis in varicose veins in the knee region that were compressed by the upper edge of the stocking in one trial.

August 2008 – SUPPORT Summary of a systematic review

Do lay health workers in primary and community health care improve maternal and child health?

Lay health workers have no formal professional education, but they are usually provided with job-related training. They can be involved in either paid or voluntary care. They perform diverse functions related to health care delivery and a range of terms are used to describe them including village health workers, community volunteers and peer counsellors among others.

Key messages

➔ **The use of lay health workers in maternal and child health programmes shows promising benefits compared to usual care or no intervention in:**
- increasing the uptake of immunization in children;
- promoting breastfeeding;
- reducing mortality in children under five years;
- reducing morbidity from common childhood illnesses.

➔ **Little evidence is available regarding the effectiveness of substituting health professionals for lay health workers or the effectiveness of alternative strategies for training, supporting and sustaining lay health workers.**

➔ **Factors that need to be considered to assess whether the intervention effects are likely to be transferable to other settings include:**
- financial support for lay health worker programmes;
- the availability of routine data on who might benefit from the intervention (e.g. children whose immunization is not up-to-date);
- resources to provide clinical and managerial support for lay health workers;
- the availability of drugs.

Who is this summary for?

People making decisions concerning use of lay health workers in primary and community health care.

! This summary includes:

- **Key findings** from research based on a systematic review
- **Considerations about the relevance of this research** for low and middle-income countries

X Not included:

- Recommendations
- Additional evidence not included in the systematic review
- Detailed descriptions of interventions or their implementation

This summary is based on the following systematic review:

Lewin SA, Babigumira SM, Bosch-Capblanch X, Aja G, van Wyk B, Glenton C, Scheel I, Zwarenstein M, Daniels K. Lay health workers in primary and community health care: A systematic review of trials, 2006. www.who.int/rpc/meetings/LHW_review.pdf

What is a systematic review?

A summary of studies addressing a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise the relevant research, and to collect and analyse data from the included studies.

SUPPORT – an international collaboration funded by the EU 6th Framework Programme to support the use of policy relevant reviews and trials to inform decisions about maternal and child health in low and middle-income countries.
www.support-collaboration.org

Glossary of terms used in this report:
www.support-collaboration.org/summaries/explanations.htm

Background references on this topic:
See back page

Background

Growing concern regarding the human resource crisis in health care has renewed interest in the roles that lay health workers may play in primary and community care delivery. This summary is based on a 2006 update of a Cochrane systematic review published in 2005 by Lewin et al. The summary focuses on the effects of lay health worker interventions in improving maternal and child health.

How this summary was prepared

After searching widely for systematic reviews that can help inform decisions about health systems, we have selected ones that provide information that is relevant to low and middle-income countries. The methods used to assess the quality of the review and to make judgements about its relevance are described here: http://www.support-collaboration.org/summaries/methods.htm

Knowing what's not known is important

A good quality review might not find any studies from low and middle-income countries or might not find any well-designed studies. Although that is disappointing, it is important to know what is not known as well as what is known.

About the systematic review underlying this summary

Review objective: To assess the effects of lay health worker (LHW) interventions in improving maternal and child health in low and middle-income countries

	What the review authors searched for	What the review authors found
Interventions	Randomised controlled trials of LHW (paid or voluntary) interventions in maternal and child health	48 trials relevant to mother and child health and high burden diseases were included and analysed
Participants	LHWs: any health worker without formal certification who was trained in some way in the context of the intervention. No restriction on types of patients	Considerable differences in numbers, recruitment methods and training of LHWs. Different recipients were targeted.
Settings	All primary care and community health settings globally	Studies from USA (25), Canada (1), UK (3), Ireland (1), Brazil (2), Mexico (1), New Zealand (1), Turkey (1), South Africa (2), Tanzania (2), Ethiopia (1), Ghana (1), Bangladesh (1), Thailand (1), Vietnam (1), India (1), Nepal (1), Pakistan (1) and the Philippines were included
Outcomes	***Primary outcomes:*** health behaviours and health care outcomes including harms. ***Secondary outcomes:*** utilization of LHW services, consultation processes, satisfaction with care, costs, social development measures	Most studies reported multiple effect measures and many did not specify a primary outcome.

Date of most recent search: August 2006

Limitations: This is a good quality systematic review with only minor limitations.

Lewin SA, Babigumira SM, Bosch-Capblanch X, Aja G, van Wyk B, Glenton C, Scheel I, Zwarenstein M, Daniels K. Lay health workers in primary and community health care: A systematic review of trials. WHO; 2006. http://www.who.int/rpc/meetings/LHW_review.pdf

Summary of findings

The review included 48 studies relevant to maternal and child health care. Most studies (26) were done in North America. Sixteen studies were conducted in low and middle-income countries in South America, Africa and Asia. Studies conducted among low-income groups in high-income countries were included in the review based on the premise that low-income groups share similar constraints in accessing health care across different countries.

1) Immunisation uptake in children under five

The six studies identified employed systems to track and remind patients whose vaccinations were not up-to-date or not vaccinated. Two studies were excluded from the meta-analyses, one study focusing on adults, and another study conducted in a very different setting.

➔ **The meta-analysis showed evidence of moderate quality that lay health worker based promotion strategies can increase immunisation uptake in children.**

About quality of evidence (GRADE)

⊕⊕⊕⊕
High: Further research is very unlikely to change our confidence in the estimate of effect.

⊕⊕⊕◯
Moderate: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

⊕⊕◯◯
Low: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

⊕◯◯◯
Very low: We are very uncertain about the estimate.

For more information, see last page.

Immunisation uptake in children under five

Patients or population: Children less than five years
Settings: Formal or informal low-income communities in USA (3 studies) and Ireland (1 study)
Intervention: Lay health worker interventions to promote immunisation uptake
Comparison: Usual care

Outcomes	Comparative risks		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)
	Without lay health workers	**With lay health workers**			
Vaccination complete according to schedule	**50 per 100**	**61 per 100** (55 to 69)	RR 1.22 (1.10 to 1.37)	3568 (4 studies)	⊕⊕⊕◯ Moderate

CI: Confidence interval RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see above and last page)

2) Mortality and morbidity in children under five years

Seven studies implemented in low and middle-income countries were identified. The main purpose of the interventions was to promote health and in some cases to manage or treat common childhood illness, including acute respiratory infections, malaria, diarrhoea and malnutrition. In four of the studies, lay health worker tasks included mainly visiting homes to educate mothers. In three of the studies, a multi-faceted package of interventions was used. Mortality and morbidity were each measured in four studies. For the mortality analysis, data from one study were excluded due to poor methodological quality. One study was excluded from the morbidity analysis as it presented insufficient data.

- **There is high quality evidence that lay health worker interventions reduce mortality in children under five years compared to usual care.**
- **There is moderate quality evidence that lay health worker interventions reduce morbidity from common illnesses in children under five years, compared to usual care.**

Mortality and morbidity in children under five

Patients or population: Children less than five years
Settings: Ethiopia, Tanzania, Nepal, Ghana, Thailand, Vietnam
Intervention: LHW interventions to reduce mortality and morbidity in children under five years of age
Comparison: Usual care

Outcomes	Comparative risks		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)
	Without lay health workers	**With lay health workers**			
Mortality	**4 per 100**	**3 per 100** (2 to 4)	RR 0.70 (0.55 to 0.99)	35,828 (3 studies)	⊕⊕⊕⊕ High
Morbidity (from fever, acute respiratory infection or diarrhoea)	**40 per 100**	**32 per 100** (28 to 36)	RR 0.81 (0.71 to 0.92)	7,544 (3 studies)	⊕⊕⊕◯ Moderate

CI: Confidence interval RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see above and last page)

3) Breastfeeding

Six studies were identified from low and middle-income countries. Lay health workers were commonly peers or volunteers selected from the community. Activities implemented included postnatal counselling to promote exclusive breast feeding and to address barriers to breastfeeding, observation of mother-child interaction, and health education.

- **Moderate quality evidence indicates that lay health worker interventions had a large effect on exclusive breastfeeding up to six months.**
- **Lay health worker interventions might increase the initiation of breastfeeding, and might slightly increase any breastfeeding up to six months.**

Breastfeeding

Patients or population: Breastfeeding mothers
Settings: Mexico, Bangladesh, Philippines and India
Intervention: Lay health worker interventions to promote initiation of breastfeeding, any breastfeeding and exclusive breastfeeding up to six months of age
Comparison: Usual care

Outcomes	Comparative risks		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)
	Without lay health workers	**With lay health workers**			
Initiated breastfeeding	**20 per 100**	**40 per 100** (16 to 98)	RR 1.98 (0.80 to 4.89)	1881 (3 studies)	⊕⊕⊕◯ Moderate
Any breastfeeding up to 6 months	**65 per 100**	**76 per 100** (64 to 91)	RR 1.17 (0.98 to 1.40)	2295 (4 studies)	⊕⊕◯◯ Low
Exclusive breastfeeding 6 weeks to 6 months	**20 per 100**	**73 per 100** (33 to 100)	RR 3.67 (1.66 to 8.11)	3021 (5 studies)	⊕⊕⊕◯ Moderate

CI: Confidence interval RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see above and last page)

Relevance of the review for low and middle-income countries

➔ Findings	▷ Interpretation*
APPLICABILITY	
➔ **The randomised trials reviewed covered an extensive range of settings, including 16 from low and middle-income countries and 26 studies from North America. The range of study settings and the consistent pattern of findings suggest that the measured effects may be transferable across settings for these health issues.**	▷ *Factors that need to be considered to assess whether the intervention effects are likely to be transferable to other settings include:* *– the availability of routine data on who might benefit from the intervention (e.g. children whose immunization is not up-to-date;)* *– resources to provide clinical and managerial support for lay health workers;* *– the availability of drugs;* *– financial support for lay health worker programmes.*
EQUITY	
➔ **Overall, the included studies provided little data regarding differential effects of the interventions for disadvantaged populations.**	▷ *Some interventions relied on technologies that may not always be appropriate when attempting to contact low-income households. Implementation of interventions in such settings utilising such technologies may exacerbate health inequities, or fail to address them adequately. Technologies employed in interventions, such as follow up via telephone calls, should be sensitive to the needs of disadvantaged populations.*
COST-EFFECTIVENESS	
➔ **The findings summarised here are based on randomised trials in which the levels of organization and support were potentially higher than those available outside of research settings. The review did not address how such support should best be provided.**	▷ *Providing adequate support to programmes is likely to be vital to intervention effectiveness when scaling up. Widespread implementation of these programmes may increase demand for services such as immunizations or treatment. If these services are not available, the activities of lay health workers may be undermined.*
➔ **Few studies reviewed described how lay health worker provided services were linked to other health system components.**	▷ *This may create difficulties when scaling up the interventions.*
➔ **Consumer participation in lay health worker programmes was also generally poorly described.**	▷ *If such participation is seen as important to programme success, considerable resources may need to be invested in this process.*
	▷ *Lay health workers are most likely to be useful when they have an effective health care intervention to deliver. Before these programmes are scaled up, robust evidence is therefore needed regarding both the effectiveness of the intervention to be delivered and of lay health workers as a delivery mechanism.*
MONITORING & EVALUATION	
➔ **Most of the lay health worker interventions shown to be effective were focused on very specific health issues. Little evidence was identified regarding the effectiveness of 'generalist' lay health workers who are given responsibility for delivering a range of primary health care interventions.**	▷ *Where lay health worker programmes are implemented for health issues for which good evidence of effectiveness is, as yet, unavailable, robust mechanisms of evaluation should be built into the programme.*
	▷ *The acceptability of lay health worker programmes to consumers and health professionals may need to be evaluated in some settings before such programmes are taken to scale.*

*Judgements made by the authors of this summary, not necessarily those of the review authors, based on the findings of the review and consultation with researchers and policymakers in low and middle-income countries. For additional details about how these judgements were made see: http://www.support-collaboration.org/summaries/methods.htm

Additional information

Related literature

Reviews evidence on the feasibility and effectiveness of community health worker (CHW) programmes in providing basic health services and addressing the shortage of health workers in low-income countries:
Lehmann U, Sanders D. Community health workers: what do we know about them? The state of the evidence on programmes, activities, costs and impact of health outcomes of using community health workers. World Health Organization, 2007.

This book summarises the findings of evaluations of large scale community health worker programmes in the 1980s, drawing out the implications for policy and practice:
Walt G. Community health workers in national programmes: just another pair of hands? Milton Keynes: Open University Press, 1990.

A systematic review of the effects of lay health worker programmes in the USA:
Swider S, M. Outcome effectiveness of community health workers: an integrative literature review. Public Health Nurs. 2002; 19:11–20.

This review summarises the evidence regarding the effectiveness of traditional birth attendants – a form of lay health workers:
Sibley LM, Sipe TA, Brown CM, Diallo MM, McNatt K, Habarta N. Traditional birth attendant training for improving health behaviours and pregnancy outcomes. The Cochrane Database of Systematic Reviews 2007, Issue 3.

This review reports earlier findings regarding the effectiveness of lay health worker interventions, including for health issues not covered in the MCH report summarised here:
Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, Bosch-Capblanch X, Patrick M. Lay health workers in primary and community health care. The Cochrane Database of Systematic Reviews 2005, Issue 1.

This summary was prepared by

Signe Flottorp, Norwegian Knowledge Centre for the Health Services, Oslo, Norway.

Conflict of interest

None declared. For details, see: http://www.support-collaboration.org/summaries/coi.htm

Acknowledgements

This summary has been peer reviewed by: Xavier Bosch, Spain; Simon Lewin, UK; Luis Gabriel, USA; Tara Bickis, Bolivia; Tracey Perez Koehlmoos, Rukhsana Gazi and Shahed Hossain, Bangladesh.

This summary should be cited as

Flottorp S. Do lay health workers in primary and community health care improve maternal and child health? A SUPPORT Summary of a systematic review. August 2008. http://www.support-collaboration.org/summaries.htm

About quality of evidence (GRADE)

The quality of the evidence is a judgement about the extent to which we can be confident that the estimates of effect are correct. These judgements are made using the GRADE system, and are provided for each outcome. The judgements are based on the type of study design (randomised trials versus observational studies), the risk of bias, the consistency of the results across studies, and the precision of the overall estimate across studies. For each outcome, the quality of the evidence is rated as high, moderate, low or very low using the definitions on page 3.

For more information about GRADE:
www.support-collaboration.org/summaries/grade.pdf

SUPPORT collaborators:

The Alliance for Health Policy and Systems Research (HPSR) is an international collaboration aiming to promote the generation and use of health policy and systems research as a means to improve the health systems of developing countries.
www.who.int/alliance-hpsr

The Cochrane Effective Practice and Organisation of Care Group (EPOC) is a Collaborative Review Group of the Cochrane Collaboration: an international organisation that aims to help people make well informed decisions about health care by preparing, maintaining and ensuring the accessibility of systematic reviews of the effects of health care interventions. www.epoc.cochrane.org

The Evidence-Informed Policy Netowrk (EVIPNet) is is an initiative to promote the use of health research in policymaking. Focusing on low and middle-income countries, EVIPNet promotes partnerships at the country level between policy-makers, researchers and civil society in order to facilitate both policy development and policy implementation through the use of the best scientific evidence available.
www.who.int/rpc/evipnet/en/

For more information, see:
www.support-collaboration.org

To receive e-mail notices of new SUPPORT summaries, go to:
www.support-collaboration.org/summaries/newsletter/

To provide feedback on this summary, go to:
http://www.support-collaboration.org/feedback/