

This is the authors' final, peer reviewed manuscript published in

Journal of Clinical Epidemiology 2010 Jun;**63(6):607-19** with the same title.

<http://www.jclinepi.com/article/S0895-4356%2810%2900024-7/abstract>

User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews

Sarah E Rosenbaum^{1*}, Claire Glenton², Hilde Kari Nylund, Andrew D Oxman¹

¹Norwegian Knowledge Centre for the Health Services

PO Box 7004, St Olavs plass

N-0130 Oslo, Norway

²Department of Global Health, Sintef, Oslo Norway

*Corresponding author. Email: sarah@rosenbaum.no

Accepted for publication in

Journal of Clinical Epidemiology

22 December 2009

Published: May 2010

ABSTRACT

Objective: To develop a Summary of Findings table for use in Cochrane reviews that is understandable and useful for health professionals, acceptable to Cochrane Collaboration stakeholders and feasible to implement.

Study Design and setting: We gathered stakeholder feedback on the format and content of a Summary of Findings table from an advisory group of over 50 participants and their constituencies through email consultations. We conducted user tests employing a think-aloud protocol method, collecting feedback from 21 health professionals and researchers in Norway and the UK. We analysed the feedback, defined problem areas and generated new solutions in brainstorming workshops.

Results: Stakeholders were concerned about precision in the data representation and about production feasibility. User testing revealed unexpected comprehension problems, mainly confusion about what the different numbers referred to (class reference). Resolving the tension between achieving table precision and table simplicity became the main focus of the working group.

Conclusion: User testing led to a table more useful and understandable for clinical audiences. We arrived at a Summary of Findings table that was acceptable to the stakeholders and in principle feasible to implement technically. Some challenges remain, including presenting continuous outcomes and technical/editorial implementation.

Keywords

Knowledge translation; health numeracy; risk communication; systematic reviews; information design; usability; user experience; fuzzy traces theory

What's new?

Key finding: We have developed a Summary of Findings (SoF) table for presenting results from systematic reviews that strikes a balance between precision and simplicity.

What this adds to what is known? How results are presented in SoF tables (including details about numerical representation, text and visual formatting) strongly influence users' perceptions and understanding of the data.

What are the implications, what should change now? Numbers in a table that need to be compared should belong to the same class. All numbers should be labelled explicitly so that class reference becomes apparent. Enabling easy gist extraction may also make the table less error prone.

“Everything should be made as simple as possible but not simpler.” (*Albert Einstein*)

“Simplicity is highly overrated” (*Donald Normann*)

BACKGROUND

Limited time is a frequently cited barrier to clinicians’ use of evidence in practice[1-6]. Systematic reviews help to address this problem by summarising evidence[7], but are still too time consuming to be practical for busy professionals. Further summarization of systematic reviews could help make evidence more useful and easy to find for decision makers with limited time[8, 9].

This is the first of two articles on the development and evaluation of summaries of Cochrane reviews for clinicians and other typical users of The Cochrane Library or general medical journals[10]. The challenges and solutions we discuss here are also likely to be relevant for other systematic reviews and health technology assessments.

Summarized evidence for clinicians exists in many different formats, for instance as structured abstracts, synopses published in secondary journals and online services. Hayne’s 5-S pyramid describes a typology of increasingly condensed and clinically useful formats: from studies (and their abstracts) to syntheses (systematic reviews), synopses (e.g. ACP Journal Club), summaries (e.g. Clinical Evidence) and systems (e.g. EPJ reminders)[11-14]. The PRISMA statement provides a consensus-based checklist for producing abstracts for systematic reviews [15], and Hartley reviews how these abstracts might be made clearer for a wide target group [16]. However little research has been published describing how clinicians experience summaries of systematic reviews. Numerical presentations of risk can be difficult, even for highly educated populations.[17] On the other hand, risk communication studies have shown that text-based descriptions of the effect of an intervention tend to be interpreted inconsistently by different people[18-20] and that numbers may be preferred by people making important health care decisions[21].

Earlier work on creating summaries of Cochrane reviews has also illustrated that to summarize already synthesized evidence is challenging and can lead to misrepresentation of the original data[22]. When attempting to summarize evidence for consumers with back pain, researchers encountered several

critical issues:

- Large numbers of reported outcomes made it difficult to identify those outcomes that are most clinically relevant
- Critical information was missing, e.g. information about adverse effects and scales
- Lack of standardization in the numerical presentation of results, the qualitative description of these results, and the manner in which the quality of data was evaluated made understanding difficult

The GRADE system offers possible solutions to some of these challenges. GRADE is a structured, transparent system that allows authors to evaluate and report the quality of evidence[23, 24]. An output of GRADE is a “Summary of Findings” table, where authors are encouraged to focus on the most important outcomes, including those outcomes with no data or statistically non-significant data and adverse effects. Authors’ judgements about the quality of evidence are presented together with the results for each outcome. The GRADE Summary of Findings table offers a useful starting point for summary authors by bringing the most important information to the foreground, regardless of the results or lack of them, and explicitly highlighting the quality of the evidence for each outcome.

Since 2004, open discussions have taken place in the Cochrane Collaboration about including Summary of Findings tables in Cochrane reviews[25], and extensive input has been gathered from stakeholders on the content and formatting of such tables. However, a number of issues continued to remain unresolved. A working group was therefore established to continue developing a Summary of Findings (SoF) table designed for inclusion in Cochrane reviews and to evaluate this table.

The SoF table should summarize the key results of the review by presenting what is known and not known about the benefits and harms of an intervention, as well as how sure we can be of the evidence. It should be understandable and useful for a clinical audience, without oversimplifying or incorrectly presenting the data. We also needed to ensure that the content and data presentation was acceptable to Cochrane stakeholders and that the formatting was feasible to produce within the technical constraints of the system for publishing Cochrane reviews. In this article, we present and discuss the development process that led to our final decisions regarding table content, format and data representation. In a

second article[10], we present the effect of including a table in a Cochrane review on user satisfaction, understanding and time spent finding key results.

METHODS

In order to develop a table that works for different types of data, we searched for a Cochrane review that included dichotomous and continuous outcomes and outcomes with no data. The Cochrane review on the effect of compression stockings for preventing deep vein thrombosis in airline passengers[26] had all of these types of results. It also covered a topic that was of potential interest to many people, making it easy to use in an evaluation process involving participants with different backgrounds. Using GRADE, we generated a SoF table for this review (Figure 1: SoF table version 0).

We used cycles of multiple methods to develop the table:

- Advisory group feedback to inform table development from a stakeholder perspective
- User testing methods to inform table development from a user perspective
- Brainstorming workshops to generate ideas and solutions to problems uncovered through feedback and testing

We also carried out two RCTs between development cycles to measure user satisfaction, correct understanding and time spent to find main messages in the review, the results of which are reported in another article [10]. We fed all stakeholder and user feedback into the brainstorming workshops. For an overview of the entire process, see Figure 2: Flow Chart.

Brainstorming workshops

We began the project with a brainstorming workshop where a working group of four people met to *generate a range of ideas* to address the issues uncovered by the other methods. We applied principles from our professional perspectives including information design, journalism and clinical epidemiology. Workshops were repeated after each round of advisory group feedback and user testing.

Advisory group feedback

The advisory group *provided feedback on the table from a stakeholder perspective*. This group consisted of over 50 people with a range of roles in the Cochrane Collaboration, including statisticians and other

methodologists, review authors, editors, consumer representatives, publishers, and members of the Steering Group. We consulted them by e-mail at three different phases of the development, encouraging them to collect feedback from their constituencies when reporting back to us. We then analysed their feedback, looking for issues with a high level of agreement or disagreement, issues we had not previously considered, or issues of critical importance such as incorrect presentation of data or formatting that was not technically feasible.

User testing

User testing provided *feedback from a user perspective*. Participants from Norway and the UK with a variety of healthcare related professional backgrounds took part in these tests.

Participants

For the first set of user tests we recruited participants attending a Norwegian workshop for newcomers to evidence-based practice. Workshop leaders asked for volunteers who could describe the basic principles of a systematic review and who had visited the Cochrane Library at least once, to minimize confounding due to unfamiliarity with Cochrane or systematic reviews. Participants' backgrounds were primarily clinical, and English was not their first language. For the second set of tests we recruited participants through the Centre for Evidence-Based Medicine in Oxford, UK. Potential participants were identified by the Centre, who contacted them by telephone or email. Though we used the same inclusion criteria as above, this group was on the whole more familiar with Cochrane reviews. Although it included several clinicians, many had a more research-related background than the Norwegian participants. English was the first language of all members of this group.

Think aloud protocol

The user tests were performed individually and took one hour. With the participant's written permission, we audio-recorded each test, and an observer took notes. Using a semi-structured interview guide, we explored immediate first impressions of the table as a whole and then detailed descriptions of each table element. The interview guide was designed to explore six of the seven different facets of "user experience" as described in a model by Peter Morville[27]: usability (defined here as "correct understanding and ease of use"), credibility, usefulness, desirability, findability and value. (See figure 4.)

The seventh facet of this model – accessibility – was not addressed as we were still testing on paper and could not explore issues relevant to online accessibility. Follow-up questions covered overall impressions and suggestions for improvement.

User test data analysis

One designer and one researcher reviewed all of the notes and transcriptions together, looking for barriers and facilitators to the six facets referred to above and tracing findings back to the elements or characteristics of the tables that appeared to cause problems. Findings were rated in three categories according to the severity of the problem: *high* (critical errors such as incorrect interpretation or high degree of uncertainty or dissatisfaction), *medium* (much frustration or unnecessarily slow use), and *low* (minor or cosmetic problems). We also registered things users explicitly liked and suggestions for improvement.

These findings were discussed in the brainstorming workshops, particularly those of high severity. For some issues, specific input was sought from the advisory group.

RESULTS

Brainstorming workshop results

In the workshops we initially focused much of our efforts on improving legibility and comprehension through changes in visual and verbal elements. For instance, to highlight key information while taking into account the technical constraints of the publishing system, we made the following changes:

- Reordered the data columns (placing results first to make them easier to locate)
- Deleted all vertical lines to emphasize horizontal reading of the rows
- Used narrower font and moved some content to the table footnotes to make the table less overwhelming in size
- Created visual “layering” of the data through use of different weights and sizes of type and use of background cell colour so that some elements visually popped forward and others fell into the background.

We made continual efforts to find terms and phrases that correctly described the data but that could be understood by non-statisticians. We initiated an explanation sheet for descriptions of terms used in the table (Table 1: Explanations for Cochrane Summary of Findings tables).

As we collected input from the advisory group and the user tests, the main focus in the brainstorming workshops became more apparent: to address the tension between achieving precision and simplicity. Tables that included enough information to meet the precision goals of the advisory group tended to be too complicated for participants to understand or want to read. There was therefore a continuous re-evaluation about what information was most critical to include and much effort was spent trying to find solutions that accommodated both perspectives.

Advisory group feedback results

We received 58 responses from 52 individuals or groups. Comments fell mostly into two categories: precision of the data representation and feasibility of publishing the tables within the current Cochrane system.

In general, the advisory group was concerned with presenting information in a form that they thought users would understand. However, there was some resistance to taking this too far:

- *“We should be extremely cautious about simplifying things to aid peoples' perception of what they are understanding”*
- *“surely even the least quantitative users will know whether 1/1000 is smaller than 10/1000, and anyone who doesn't should not be allowed to use the findings of a Cochrane review!”*

Feedback related to precision of data representation included comments about:

- missing data, for instance:
 - *“We need to know the duration for the effect, in this case it's per flight: >6 hours in duration.”*
 - *“It should be mandatory to explain the basis for the assumed control group risk...”*
 - *“All the reasons for the quality being limited should be described in the footnotes”*

inaccurate or potentially misleading elements, for instance

- *“I would suggest... omitting 'favours intervention' and 'favours control. (T)he statement 'Favours X' is arguably misleading because (...) for some outcomes it is unclear whether a reduction in risk is good or bad, and you may encourage review authors to impose their subjective judgment”*
- *“Ideally there should be some recognition of imprecision about the rates/values in the control group - the impact of not allowing this is that differences in absolute values are artificially precise.”*

Examples of feedback regarding production and publishing within the Cochrane system:

- *“I was very skeptical about your ability to make the multiple control group risks understandable, but it looks to me as if you've done it with the variations in cell color and in fonts. Now the next hurdle is to find a way to actually get the published tables to look like your example.”*
- *“My main concern is the roll-out of changes to Cochrane reviews (like SoF Tables) balancing the need for development with the challenges of making changes to hundreds of reviews”*

User testing results

Twenty one people from Norway and UK took part in the user tests. During the first set of tests we found several problems that we ranked as *high severity*. After modifying the table several times, we tested a new version. No findings in the *high severity* category were observed in this second set of user tests. The findings that led to most changes in the table were concentrated in two of the seven facets of the user-experience model: *usability* and *usefulness*.

Usability (correct understanding and ease of use)

A major finding, particularly in the first set of user tests, was that participants misunderstood or were uncertain about a range of elements:

- dichotomous outcomes
- continuous outcomes

- number of studies
- meaning of “no data available” or empty cells
- terms used in column headings
- abbreviations

For instance, five of 13 participants dramatically misunderstood “9 fewer per 1000” in the column for “Absolute difference”, stating that it meant “9” or “9 or fewer”. This mistake was made by some even when they correctly read the effect statement out loud. Two participants understood the statement correctly but were unsure if their interpretation was right. Three of 13 participants mentioned specifically that they used *"Favours stockings"* to confirm that they had understood the numbers correctly.

Continuous outcomes caused confusion, usually because participants could not identify what the numbers related to: *"5 to 9 what? People?"* Explanations, placed in the Comments column, were often overlooked. Other numbers also caused confusion: four of 13 test persons in the first set of user tests said that the number of studies “(9)” was either a reference to a footnote or they did not know what it meant.

Participants also exhibited unfamiliarity with language and concepts used in the table. Sixteen of 21 participants did not understand the headings “Illustrative comparative risk”, “Assumed risk” and “Corresponding risk” and 12 of 21 did not understand what was meant by “no data available” or empty cells. Abbreviations such as “RR” (relative risk) and “CI” (confidence interval) also caused confusion regarding both what the abbreviation stood for and the concept it referred to.

Participants did not have critical problems related to understanding the GRADE ratings, despite most not having prior knowledge of GRADE.

Usefulness

Participants offered suggestions for changes that would make the tables more useful in a clinical setting. These included:

- Specifying the population, setting, intervention and control group at the top of the table

- Describing the intervention in more detail
- Adding the inclusion criteria for high and low risk populations
- Including a clear recommendation

Credibility

Eighteen of 21 test persons indicated that their perception of the credibility of the table was directly related to the GRADE ratings. *“I would say that if the quality of evidence (referring to the GRADE score) was high, then I would believe in it more”.*

Findability

Most participants indicated that a Summary of Findings table should be near the front of the review, near the abstract. User preference regarding placement was measured explicitly in our randomized trial of the table[10].

Desirability and value

Fourteen of 21 participants said that the table would be a valuable addition to Cochrane reviews. One person did not like tables in general. One participant explained she did not like it but anticipated that she would feel differently over time after becoming more familiar with the format. User satisfaction was also measured in our randomized trial[10].

First impressions versus exposure over time

Although 11 of 21 participants felt the table contained large amounts of information, this was not necessarily negative. Some said they expected a learning curve for this kind of information and were confident that they would find these tables easier to read upon repeated exposure.

- *“...I spent a lot of time but when I first broke the code I found it easy... next time it will be better.”*
- *“Immediate reaction (was) oh lots of figures, lots of numbers, but after a minute...when I go systematically...its sort of quite good. The more I look the more I like it”*
- *“(My first impression is that it is) a big table with a lot of information... but I’m not de-*

motivated because I think that there is something credible here.”

Resulting SoF table

Our work resulted in many iterations of the SoF table. Figure 3 shows the last version.

DISCUSSION

Through feedback from the advisory group and our efforts in the brainstorming workshops, we arrived at a table that was acceptable to the stakeholders and in principle feasible to implement technically. User testing helped us to improve the table for a clinical audience. There are remaining challenges, including presenting continuous outcomes and implementing the table in the Cochrane publishing system.

Prior to the start of our project, the GRADE Working Group had made several choices regarding the formatting of the table guided by what was known about how people understand risk information. One key choice was that data should be represented numerically, partly because this would provide a supplement to the already text-based abstract and plain language summary, but also because a numerical presentation of results would be a more precise starting point for other summaries based on the review.

The manner in which numerical results were presented was also guided by research evidence indicating that:

- Absolute risk (including baseline rates) should be presented as well as relative risk[28].
- NNT (numbers needed to treat) and NNH (numbers needed to harm) are difficult when there are multiple outcomes or statistically non-significant effects.
- Event rates (1 out of 1000) may be easier to understand than percentages, because they help identify the reference class in question.[29, 30]
- Denominators with the base of 10 (e.g. 10, 100, 1000) are easier to comprehend[18]
- Use of same denominator facilitates comparison[31]
- Symbols may be an effective format for communicating quality of evidence[32, 33]

Trouble understanding the class references

Although numbers may be more precise than qualitative presentations, they still have problems. We initially thought that the focus of our project was to arrive at a table that users were satisfied with. However, achieving user satisfaction does not guarantee that information is being understood *correctly*. During the first set of user tests we became aware that correct comprehension was a much larger issue than we had anticipated. Much of the difficulty that we observed was related to confusion about what numbers referred to (“class reference”). Problems correctly identifying reference class have been uncovered in past work[30, 34].

Trouble with Absolute Effect

Instead of making the table easier to read by reducing computational tasks, the statement “9 fewer per 1000” caused uncertainty and errors. This is possibly due to the subtle change of class reference between the control group risk column and absolute effect column: “X number of people per 1000” and “X fewer number of people per 1000”. In a recent review of formats for conveying health risks, Lipkus recommends consistency in use of numerical formats [18]. When we reformatted the way magnitude of effect was represented in this column - eliminating the Absolute Difference format (x fewer per 1000) and changing it to Absolute Risk (x per 1000) - users no longer made these errors.

Continuous outcomes – continuous challenge

Many test participants also struggled to interpret continuous outcomes. This problem also seemed to be related to inconsistent class references: dichotomous results and continuous results appeared in the same columns, but the numbers for these two outcome types referred to different classes of phenomena. “1 per 1000” refers to numbers of people while “mean 6 to 9” refer to a range on a scale. We experimented presenting continuous outcomes using both sentences and numbers so that the scale references became more apparent, but are uncertain how effective this format is as it was not tested explicitly.

In addition, the column heading “Corresponding Risk With Stockings” is technically wrong for these outcomes. This kind of discrepancy could be dealt with if the text in column headings were less precise, for instance only “Without Stockings” and “With Stockings”, leaving the more accurate descriptions of the column content to a footnote. This issue and the issue of how to present continuous outcomes needs further work.

Trouble identifying other numbers' class references

Readers' uncertainty about the class reference also cropped up in other places. Throughout the table, different numbers refer to different classes of things. Figure 5 (an early version of the table) illustrates this more clearly. Here "30/1000" in the DVT row refers to people, "(1 to 8)" refers to per 1000 people, "(8)" refers to studies, whereas "6 to 9" in the oedema row refers to range on a scale. Although the row and column headings explain what these different numbers mean, this was not enough for many participants. When the formatting is similar but means two different things, such as "6 to 9" meaning range on a continuous outcome scale and "(1 to 8)" meaning confidence interval, readers at any level may be challenged.

Trade-offs between class cues and clutter

Difficulties associated with class reference has been pointed out in earlier studies: combining information from different classes, leaving class open to interpretation[30] and overlapping or nested classes[35]. The confusion we observed appeared to be due to difficulty identifying different classes. Text labels in direct proximity to the numbers (e.g. "Mean oedema range: 6 to 8" or "9 studies") help clarify the class reference. The trade-off is to balance this information without creating an overly cluttered table that may both demotivate readers and also interfere with their task of quickly taking in key information.

Precision or simplicity - Verbatim or gist?

The tug-of-war between precision and simplicity, reflected in the feedback from the advisory group and the test participants, was our main challenge when designing the table. A good example of this conflict was the differing feedback we received concerning the phrase "Favours stockings". This phrase was inserted directly underneath the numbers expressing magnitude of effect for one outcome in an early version of the table. User test participants said that this phrase was helpful, explaining that this kind of cue helped them feel more confident in their understanding of the figures. The advisory group recommended taking these phrases out, because they were seen as misleading and oversimplifying. However, participants' favourable reactions to "Favours stockings" may tell us something about how numbers are actually used in decision making.

There is some evidence suggesting that people may not calculate with precise numbers (such as "10 per

1000” or “1 per 1000”) in real-world problem situations involving decision making or interpretation, but prefer to rely on gists - semantic representations of the information)[36, 37] - such as “Favours stockings”. Fuzzy traces theory can explain this preference, claiming that people display a dual-processing of information along a verbatim-gist continuum. Readers register both the verbatim (the precise information) and the gist (the qualitative interpretation of what is being communicated), but have a gist-preference[38]. Extracting the correct gist can prevent basic comprehension errors[35]. Some work also indicates that gist-preference may increase with higher levels of expertise[39-41].

“Favours stockings” is a phrase that helps readers quickly form a correct gist of what the numbers mean, and saves time. This gist may be sufficient to answer a decision maker’s initial questions about a particular outcome - “Did the intervention have an effect? (yes/no)” and “Is this effect desirable in this situation? (yes/no)” - before actually paying attention to the exact amount of the effect. Such levels of precision may not be necessary until both these initial questions are answered affirmatively and the process of balancing actual amounts of benefits, harms, costs and uncertainty can begin.

Perhaps equally important, the phrase “Favours Stockings” is less prone to being dramatically misunderstood (i.e. is not easily confused with “does not favour stockings”), whereas small easy-to-make mistakes in processing the precise data could produce major errors. For instance, assuming that the intervention group results were in the first of the two effect columns, or misunderstanding the framing of a continuous outcome scale (is high on this scale good or bad?), would provide a totally incorrect gist of the data. Preferences for text cues such as “Favours stockings” might reflect an appropriate safeguarding behaviour for those who feel they may be at risk of making mistakes when faced with a complex table of numbers.

If gists can not be represented explicitly, make them easy to extract

As the advisory group pointed out, although potentially helpful, the phrase “Favours stocking” may lead to over-interpretation when the effect difference is actually very small or the confidence interval is wide. The word “favours” also implies an imbedded value judgment about the desirability of the outcome that should not be made by a systematic review author[42]. Despite user preference, these cues were therefore eliminated. An alternative to providing cues may be to ensure that data is presented in ways that enables readers to easily extract the correct gist out of the verbatim information. For instance the

information can be visually layered through use of colour or varying type size/weight, so that key messages pop out more clearly [43]. Numbers can be aligned to create more visual order, aiding comparison and gist extraction. Neglecting to do so may scramble the information and render it less useable/useful as well more error-prone[34].

Technical barriers to enabling gist extraction

The table was designed to fit within the constraints of the Cochrane publishing system, though actual implementation of several features of the table have proven to be difficult, both in HTML and PDF versions. These include the features that help readers quickly focus attention on the main messages and aid gist extraction (shading of cells, variation of font type/size/weight). We are currently working to resolve these issues.

Evidence into practice – making information useful for clinical contexts

Part of the challenge of bringing research into practice is making the information useful for a clinical context. Through user testing we collected feedback on specific elements that would render the SoF table more useful in a clinical context, including specifying the criteria for high and low risk populations and describing the intervention in more detail. Glasziou has pointed out that detailed description of the intervention is critical for the clinical reader, but is often lacking in both systematic reviews and articles reporting on clinical trials[44].

Limitations

The strengths of this study include the use of multiple methods and involvement of a range of stakeholders with complementary perspectives. However, the study has some limitations:

- Participants in the second set of user tests had on average a more research-oriented background than the first group. Therefore the lack of critical problems in the second set of tests may not be representative.
- The use of the table was not evaluated in real-life settings.
- The developers of the tables carried out the user tests, and participants were aware of this.

CONCLUSION, GUIDELINES AND FURTHER RESEARCH

Aspects of SoF table design (including details about numerical representation, text and visual formatting) have a strong influence on users' perceptions, especially regarding their understanding of the data. General guidelines for these kinds of tables are:

- Avoid class confusion:
 - use same class reference, especially in number sets that are to be compared
 - Support correct class interpretation by adding class labels (e.g. “studies”)
 - Describe scales for continuous outcomes in close proximity to the results
- Avoid unfamiliar abbreviations wherever possible, even if they have been introduced in the text.
- Explain empty cells to make uncertainty or lack of data explicit
- Help the reader quickly form the correct gist of the numbers:
 - use text cues where applicable
 - align type to make comparison of numbers easier
 - Layer the information visually so that the most important parts “pop out” at the reader

To make tables more useful for clinicians, include:

- information about the population and setting
- inclusion criteria for the high/low risk populations
- description of the intervention

The table met with broad approval by the advisory group and by the health professionals in the user testing. The Cochrane Collaboration now recommends including Summary of Findings tables in Cochrane reviews, placed after the abstract[45]. Formatting will be somewhat limited due to technical issues in the publishing system. Results from two RCT's measuring the table's effect on user satisfaction, understanding and time spent finding results in a systematic review are reported in a separate article[10].

Further work in progress includes how to update existing reviews with Summary of Findings tables, how to implement them in the production of new reviews, how to present continuous outcomes, and how to n

produce tables targeted at consumers and at policy makers. The Summary of Findings format was developed using only one example (compression stockings). Although this summary was complicated and most summaries will be simpler, other reviews may present additional challenges, such as summarizing several comparisons and presenting results for outcomes when a meta-analysis was not possible.

Future research should include comparisons of this summary table with other summary formats currently in use.

The proposed format is being used by other organizations publishing summaries of findings. Software is available to generate SoF tables using this format[46].

Acknowledgments

Thanks to Arild Bjørndal for his help with the manuscript.

REFERENCES

Figures and table legends

Figure 1: Summary of Findings table - Version 0

Figure 2: Flow chart

Figure 3: Summary of Findings table - final version

Figure 4: User Experience Model (Peter Morville)

Figure 5: From an early table version

Table 1: Explanations for Summary of Findings tables