PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A usability study of two formats of a shortened systematic review for clinicians
AUTHORS	Perrier, Laure; Kealey, M.; Straus, Sharon

VERSION 1 - REVIEW

REVIEWER	Hilda Bastian National Center for Biotechnology Information, National Institutes of Health
REVIEW RETURNED	27-Aug-2014

GENERAL COMMENTS	This is an excellent set of work, and an excellent paper, as well. However, while I checked "yes" to question 11, but the answer is "no" for the section on "strengths & limitations of this study." There is no adequate discussion or literature assessment within the paper to justify the statement in the first bullet point. Work from key groups working in this field isn't referenced, for example: http://www.ncbi.nlm.nih.gov/pubmed/23302501 http://www.ncbi.nlm.nih.gov/pubmed/20434023 http://www.ncbi.nlm.nih.gov/pubmed/20434024 http://www.ncbi.nlm.nih.gov/pubmed/21346891 http://www.ncbi.nlm.nih.gov/pubmed/24293571 That's not to say that the work here is not more rigorous in some aspects than other work in the field: just that it under-estimates the importance of other work. Two particular issues arise here: one is the key issue of randomized trials or other solid means of comparing the work with other ways of communicating similar information; and the other is generalizability. The other weak area is the section on study limitations. The group here may well have been representative of a particular group of end users in a particular cultural setting: that doesn't mean the results are generalizable outside that context. The level of expertise of the participants isn't clear from the study: they may well be far more used to systematic reviews than practitioners from other settings, would be my guess, but I can't tell from the information provided. The process was designed to see if the summaries achieved their intended results. Whether there is a better way to improve comprehension would need to be addressed in other work.

	Finally, in respect of limitations: format is not the only issue here. There are components embedded here such as editorial skill, the value of the named person providing authority to the summary (who certainly deserves to be widely trusted, and I hope she is!), and the fact that it relates to a single systematic review in a specific topic area (and we don't know how much salience that topic had for the participants). That last is a major limitation: some topics would lend themselves more readily to this format than others; the availability & nature of the evidence base varies considerably from topic to topic and that may also affect the usefulness of the format. This is not meant to under-estimate the value of this extremely solid piece of work: it's just to put it in context. That first bullet point and that one section are the only problems I have with this article. I'd just like to underscore my compliments to the people involved in this project: it's a very valuable contribution to the field and was a pleasure to read as well. Thank you

REVIEWER	Isomi Miake-Lye Veterans Affairs Greater Los Angeles Healthcare System UCLA Fielding School of Public Health USA
REVIEW RETURNED	04-Oct-2014

GENERAL COMMENTS	One general question is why there are two formats being tested. Do they serve different purposes, or is the intent to have one format in the end? Another general question: how would a reader determine the quality of a systematic review in this shorter format? What were your findings from question 13 in your interview guide (Appendix A)? These might be helpful to present. In the abstract it would be helpful to say how well the issues were resolved or whether the prototypes seem to be useful to give a sense of a take-away message, since it is unclear if the prototypes are now ready for use or need more testing. Bullets one and three in the "Strengths and limitations of this study" section could be revised. Are these bullets referring to strengths of the study? Having a bullet in this same section saying how well the prototypes performed, and which one performed better, would be helpful. Line 17, p. 8 would fit better in the area about ethics. What was the response rate during recruitment? How many participants were initial contacts vs. snowball sampling? In the Methods, please try to clarify that each participant was involved in one testing instance, and that this instance included both prototypes. This becomes clear in the results section, but the language in the paragraph starting on line 19, p. 5 is a little bit vague. Good discussion of the results, it might be helpful to add proportions or percentages to Table 2 because the sample sizes vary across cycles and comparing numbers is difficult. The Discussion and Conclusion read a little too much like a summary of results. Including more implications, future work, or potential use for this format would be more engaging. For instance,

	do you intend to continue with two formats?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 Comment
This is an excellent set of work, and an excellent paper, as well.

However, while I checked "yes" to question 11, but the answer is "no" for the section on "strengths & limitations of this study." There is no adequate discussion or literature assessment within the paper to justify the statement in the first bullet point. Work from key groups working in this field isn't referenced, for example:

http://www.ncbi.nlm.nih.gov/pubmed/23302501
http://www.ncbi.nlm.nih.gov/pubmed/20434023
http://www.ncbi.nlm.nih.gov/pubmed/20434024
http://www.ncbi.nlm.nih.gov/pubmed/21346891
http://www.ncbi.nlm.nih.gov/pubmed/24293571

That's not to say that the work here is not more rigorous in some aspects than other work in the field: just that it under-estimates the importance of other work.

Two particular issues arise here: one is the key issue of randomized trials or other solid means of comparing the work with other ways of communicating similar information; and the other is generalizability.

Author Response
Completed.
Page 3-4: The Background was enhanced to identify relevant literature focusing specifically on rigorous evaluation of information tools related to shortened formats of systematic reviews: "Numerous tools are available to clinicians that present summarized evidence-based information……."

Please note with regards to the suggested references,
• Rosenbaum (2010): 2 studies are included and discussed in this manuscript in the Background (page 3-4)
• Rosenbaum (2011) – is not relevant as specifically targeted to policymakers who are a distinct group from clinicians making this reference unrelated
• Treweek (2013) – A plan is outlined to create a tool/resource but it does not exist yet. Thus, there is no rigorous evaluation of a concrete tool.
• Rader (2014) – A description of the process used to develop tools however, a rigorous evaluation of their tool(s) is not reported. A search was completed to identify if this group had published another article reporting an evaluation but none was found.

Reviewer 1 Comment
The other weak area is the section on study limitations. The group here may well have been representative of a particular group of end users in a particular cultural setting: that doesn't mean the results are generalizable outside that context.

Author Response
Completed.
Page 15-16 – Generalizability is acknowledged in the limitations and more information added with regards to participants: "…the demographics of the sample provide some indication of a diverse

group…”

Reviewer 1 Comment
The level of expertise of the participants isn't clear from the study: they may well be far more used to systematic reviews than practitioners from other settings, would be my guess, but I can't tell from the information provided.

Author Response
Completed.
Page 16 – Level of expertise is addressed:
“Training in critical appraisal or experience in conducting systematic reviews was not collected from participants which may have indicated levels of expertise related to evidence-based information tools. However, this data would not have great impact with regards to the usability testing as problems encountered with the use of the tool was being assessed rather than the content or comprehension.”

Reviewer 1 Comment
The process was designed to see if the summaries achieved their intended results. Whether there is a better way to improve comprehension would need to be addressed in other work.

Author Response
Completed.
Clarity is provided with regards to the purpose of usability testing and explicitly stating there is no attempt to assess comprehension:
• Page 5 – referring to usability testing: “It does not test the comprehension of the content but rather provides direct information about how people use a tool and what their exact problems are with the tool being tested.”

• Page 16 – “Training in critical appraisal or experience in conducting systematic reviews was not collected from participants which may have indicated levels of expertise related to evidence-based information tools. However, this data would not have great impact with regards to the usability testing as problems encountered with the use of the tool was being assessed rather than the content or comprehension.”

Reviewer 1 Comment
Finally, in respect of limitations: format is not the only issue here. There are components embedded here such as editorial skill, the value of the named person providing authority to the summary (who certainly deserves to be widely trusted, and I hope she is!)

Author Response
Completed.
Clarity is provided with regards to the purpose of usability testing and explicitly stating there is no attempt to assess content (which would include issues related to the editorial skill of the content, credibility of the authority to the summary, etc.):
• Page 5 – referring to usability testing: “It does not test the comprehension of the content but rather provides direct information about how people use a tool and what their exact problems are with the tool being tested.”

• Page 16 – “Training in critical appraisal or experience in conducting systematic reviews was not collected from participants which may have indicated levels of expertise related to evidence-based information tools. However, this data would not have great impact with regards to the usability testing as problems encountered with the use of the tool was being assessed rather than the content or comprehension.”

Reviewer 1 Comment
…. and the fact that it relates to a single systematic review in a specific topic area (and we don't know how much salience that topic had for the participants).

Author Response
Completed.
The salience of the topic of the systematic review is described:
• Page 4-5 – “We chose a full-length systematic review to be used for the development of the prototypes from a list of recently published systematic reviews supplied by the Health Information Unit at McMaster University….”

Reviewer 1 Comment
That last is a major limitation: some topics would lend themselves more readily to this format than others; the availability & nature of the evidence base varies considerably from topic to topic and that may also affect the usefulness of the format.

This is not meant to under-estimate the value of this extremely solid piece of work: it's just to put it in context. That first bullet point and that one section are the only problems I have with this article.

I'd just like to underscore my compliments to the people involved in this project: it's a very valuable contribution to the field and was a pleasure to read as well. Thank you!

Author Response
Completed.
Full-length systematic reviews are reported using a prescribed format (Moher BMJ 2009), thus all topics would be presented with the same components. Focus groups were conducted prior to usability testing (Perrier 2014) in order to identify which of these components were essential to clinical decision-making. This was clarified on Page 4:
“The focus groups provided a forum for clinicians to identify the essential components of a format for a shortened systematic review, including key features and content, to aid in clinical decision making”

Reviewer 2 Comment
One general question is why there are two formats being tested. Do they serve different purposes, or is the intent to have one format in the end?

Author Response
Completed.
Page 4-5 - An explanation for the development of two formats is provided:
“Support in the literature was found for the development of two shortened formats…”

Reviewer 2 Comment
Another general question: how would a reader determine the quality of a systematic review in this shorter format?

Author Response
Completed.
Clarity is provided with regards to the purpose of usability testing and explicitly stating there is no attempt to assess content (which would include issues related to the quality of the systematic review, etc.):
• Page 5 – referring to usability testing: “It does not test the comprehension of the content but rather

provides direct information about how people use a tool and what their exact problems are with the tool being tested."

• Page 16 – "Training in critical appraisal or experience in conducting systematic reviews was not collected from participants which may have indicated levels of expertise related to evidence-based information tools. However, this data would not have great impact with regards to the usability testing as problems encountered with the use of the tool was being assessed rather than the content or comprehension."

Reviewer 2 Comment
What were your findings from question 13 in your interview guide (Appendix A)? These might be helpful to present.

Author Response
No changes made.
Question 13 in the interview guide (Appendix A) asked participants: Would you like the Methods of the systematic review to be described on the shortened systematic reviews?
• All participants answered 'No' to this question. As a result there is nothing to report since no problems were identified and no changes were made to the prototypes.

Reviewer 2 Comment
In the abstract it would be helpful to say how well the issues were resolved or whether the prototypes seem to be useful to give a sense of a take-away message, since it is unclear if the prototypes are now ready for use or need more testing.

Author Response
Completed.
Page 2 - Clarification provided with the following sentence added to the Abstract: "Alterations were made in order to create finalized versions of the two shortened systematic review formats."

Reviewer 2 Comment
Bullets one and three in the "Strengths and limitations of this study" section could be revised. Are these bullets referring to strengths of the study?

Having a bullet in this same section saying how well the prototypes performed, and which one performed better, would be helpful.

Author Response
Completed.
• Bullets one and three have been removed as they provide background information and are not a strength/weakness specific to the usability testing.
• The final bullet is identified more clearly as a limitation.

Page 3 – The number of errors were reported in the Strengths and limitations segment:
"Errors were found during each of the three iterative cycles of usability testing (Cycle 1: 5 errors; Cycle 2: 8 errors; Cycle 3: 6 errors)…"

Reviewer 2 Comment
Having a bullet in this same section saying how well the prototypes performed, and which one performed better, would be helpful.

Author Response
Completed.
Page 8 – The following sentence was added:
“Since the order in which the two prototypes were presented to participants was randomized and the majority of questions (13 out of 15) were identical, no attempt was made to report which prototype had more errors as viewing the first prototype provided insight as to the types of items participants would be asked to locate for the second prototype.

Reviewer 2 Comment
Line 17, p. 8 would fit better in the area about ethics.

Author Response
Incomplete.
Unfortunately, I cannot identify the specific sentence. When I add Line Numbers to an original copy, Line 17, page 8 is Table 1 which does not relate to ethics.

If more information can be provided, I can address this more appropriately.

Reviewer 2 Comment
What was the response rate during recruitment? How many participants were initial contacts vs. snowball sampling?

Author Response
Complete.
Page 7 – The following sentence was added:
“152 recruitment emails were sent to potential participants. Six physicians were recruited with this method and four through snowball sampling giving a response rate of 7%.“

Reviewer 2 Comment
In the Methods, please try to clarify that each participant was involved in one testing instance, and that this instance included both prototypes.
This becomes clear in the results section, but the language in the paragraph starting on line 19, p. 5 is a little bit vague.

Author Response
Completed.
Page 6 – The following sentence was added:
“Three iterative cycles of usability testing was completed and physicians could participate in one cycle of testing only.”

Page 7 – Clarification was provided in the Methods that each participant was shown both shortened formats:
“Both prototypes were presented to each physician…”

Reviewer 2 Comment
Good discussion of the results, it might be helpful to add proportions or percentages to Table 2 because the sample sizes vary across cycles and comparing numbers is difficult.

Author Response
Completed.

Page 13-14 - Percentages added to Table 2 for individual tasks only.
Note that a percentage cannot be reported for Total Errors as the Table does not list all tasks presented, only those where errors occurred.

Reviewer 2 Comment
The Discussion and Conclusion read a little too much like a summary of results. Including more implications, future work, or potential use for this format would be more engaging. For instance, do you intend to continue with two formats?

Author Response
Completed.
Page 16-17 - The following has been added:
“Modifications to the prototypes were implemented based on the usability testing resulting in final versions for the two shortened formats of a systematic review. We plan to conduct a pilot study in order to assess the feasibility of a full-scale randomized controlled trial where participants will be asked to apply the evidence from either the full-length systematic review or one of the shortened formats to a patient that is presented to them in a clinical scenario”

VERSION 2 – REVIEW

REVIEWER	Isomi Miake-Lye Veterans Affairs Greater Los Angeles Healthcare System UCLA Fielding School of Public Health USA
REVIEW RETURNED	04-Nov-2014

GENERAL COMMENTS	The revisions have addressed my concerns and I enjoyed reading their work. This manuscript represents an interesting development and contribution to the field!

REVIEWER	Hilda Bastian National Center for Biotechnology Information (National Library of Medicine, National Institutes of Health), USA
REVIEW RETURNED	25-Nov-2014

GENERAL COMMENTS	Nice work - thank you!