

BMJ Open A usability study of two formats of a shortened systematic review for clinicians

Laure Perrier,¹ M Ryan Kealey,² Sharon E Straus^{3,4}

To cite: Perrier L, Kealey MR, Straus SE. A usability study of two formats of a shortened systematic review for clinicians. *BMJ Open* 2014;**4**:e005919. doi:10.1136/bmjopen-2014-005919

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2014-005919>).

Received 16 June 2014
Revised 16 October 2014
Accepted 26 November 2014



CrossMark

¹Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

²Faculty of Applied Science and Engineering, University of Toronto, Toronto, Canada

³Faculty of Medicine, University of Toronto, Toronto, Canada

⁴Keenan Research Centre, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada

Correspondence to

Laure Perrier;
l.perrier@utoronto.ca

ABSTRACT

Objective: The aim of this study was to evaluate the usability of two formats of a shortened systematic review for clinicians.

Materials and methods: Usability of the prototypes was assessed using three cycles of iterative testing. 10 participants were asked to complete tasks of locating information or items within two prototypes and 'think aloud' while being audio taped. Interviews were also audio recorded and participants completed a systematic usability scale.

Results: Revisions were made between each iteration in order to address issues identified by participants. Finding information relating to the number of studies in the meta-analysis, and locating the number of studies in the entire systematic review were revealed as areas needing attention during the usability evaluation.

Conclusions: Iterative testing combined with a multifaceted approach to usability testing offered essential insight into aspects of the prototypes that required modifications. Alterations were made in order to create finalised versions of the two shortened systematic review formats.

BACKGROUND

Systematic reviews are rigorous, comprehensive assessments of evidence that provide a synthesis of the literature.¹ Many consider systematic reviews of high quality randomised controlled trials to rank as the highest in the hierarchy of evidence for interventions² and the best source of information for making clinical decisions.³ Although systematic reviews are readily available in the published literature and through collected repositories such as the Cochrane Library, there is evidence that their use in clinical decision-making is not optimal.³⁻⁴ Searching, identifying and retrieving evidence-based resources paired with the lack of time experienced by busy clinicians is consistently identified as an obstacle to answering clinical questions.⁵⁻¹⁰ One solution to this obstacle is to create resources that use explicitly formulated methodological criteria so that information can be validated and refined in order to

Strengths and limitations of this study

- Two templates for a shortened format of a systematic review have been rigorously developed with input from end users.
- Errors were made by participants during each of the three iterative cycles of usability testing (cycle 1: five errors; cycle 2: eight errors; cycle 3: six errors) and highlighted areas of the template that required refinement, such as information on treatments that had no trials.
- A limitation is that the interplay between trait characteristics and end-user experiences could not be examined and may be revealed through the examination of personal (eg, age) or experiential factors which would be important considerations for future studies with larger populations.

be read quickly.⁹ Numerous tools are available to clinicians that present summarised evidence-based information, either from a collection of sources (eg, UpToDate) or solely from systematic reviews and presented in a condensed format (eg, BMJ PICO abridged research articles). Prior to the development of our prototypes, we completed a systematic review examining the impact of interventions for seeking, appraising and applying evidence from systematic reviews in decision-making by clinicians.¹¹ During the assessment of records, we specifically screened for studies that evaluated a different strategy for presenting a systematic review. Two trials were identified that reported promising results.¹²⁻¹³ However, on closer examination, their considerable limitations indicate that the results should be interpreted with caution due to their recruitment of a convenience sample affiliated with the Cochrane Collaboration and their use of a small sample size.¹²⁻¹³ Consequently, no studies providing a rigorous evaluation of the impact on end users were found in the literature. In an effort to address this, two shortened systematic review formats were

developed to enhance their use by clinicians.^{14 15} The next step in the development process is to conduct a usability study with the prototypes so that they can be further refined and the design can be finalised.

OBJECTIVE

To describe the usability testing of prototypes for two formats (case-based and evidence-expertise) that represent a shortened version of a full-length systematic review.

METHODS

The approach to developing two alternate formats of a shortened systematic review to modify the presentation of information has been described in previous publications.^{14 15} In brief, prototypes for two formats of a shortened systematic review were developed using an explicit, rigorous process including a mapping exercise, an initial heuristic evaluation and a clinical content review.¹⁴ Following this, a series of focus groups were conducted with primary care physicians to further refine and determine the optimal format for the presentation of information.¹⁵ 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.¹⁵ Support in the literature was found for the development of two shortened formats. The first format includes a case study to present contextualised information (case-based format), and the second format integrates evidence and clinical expertise (evidence-expertise format). The case-based format was designed to present a real-world example of how the clinical evidence would be used in decision-making. Text is easier to understand when it contains personalised elements, such as examples like case studies.^{16–20} This allows instructions and information to be embedded more succinctly and prompts readers to recall more information.^{19–21} The evidence-expertise format, guided by David Sackett's definition of evidence-based medicine, highlights the integration of clinical expertise and the best external evidence.²² An assessment of over one thousand systematic reviews showed that almost half presented results that neither supported nor refuted the interventions tested.²³ In describing preferences for the presentation of evidence, primary care physicians expressed the need for an explicit statement when evidence was absent and the need for clinical expertise to bridge the gap.²⁴ We chose a full-length systematic review to be used for the development of prototypes from a list of recently published systematic reviews supplied by the Health Information Unit at McMaster University (Canada).²⁵ The systematic review had been rated by a pool of clinicians using the McMaster PLUS scale that allows them to identify articles they believe would be important to practising primary care physicians. Those that scored six or better (of seven) were reviewed by two physicians

(one internal medicine physician and one family physician) who independently voted on the three most relevant to generalist physicians. The final review was selected by a third family physician independently. The final full-length systematic review that the prototypes were derived from for this study was van Zuuren *et al.*²⁶

The final phase in the development process is to conduct a usability study using iterative cycles of testing with primary care physicians. Usability testing focuses on how well users can learn and use a product to achieve their goals and is defined as “how effectively, efficiently, and satisfactorily a user can interact with a user interface”.²⁷ 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.

Usability testing

Participants

Primary care physicians practising full-time or part-time who were able to read and speak English were identified as eligible for participation. This group was chosen as systematic reviews summarise vast quantities of information on specialised topics which can be useful for generalist physicians. The sampling strategy for usability testing involved randomly selecting physicians practising in the Toronto, Canada area from a list available from the CPSO (College of Physicians and Surgeons of Ontario) database (representing 13 298 active family physicians and 3520 general internists). We also used snowball sampling²⁸ which relies on referrals from the initial subjects to identify additional participants. Physicians were emailed and asked to reply indicating the time and date they were available to participate. Three iterative cycles of usability testing were completed and physicians could participate in one cycle of testing only.

Study design and procedures

Participants were assured of confidentiality in reporting the results. An honorarium was provided to participants.

The usability testing of the two prototypes was run in three iterative cycles with 2–5 participants per iteration. This approach is a process of implementing a design or tool, seeking discussion and feedback and making subsequent refinements to the design or tool.²⁹ Multiple testing is supported as the goal is to improve design, not just document weaknesses and sampling as little as five users has been shown to uncover substantial problems.³⁰ After consent was obtained, participants were given instructions about the usability test. The participants were given the choice of viewing and testing the prototypes on a laptop brought by the investigator or on their own computer. Participants were given case scenarios along with the modified systematic reviews and asked to complete a task relevant to the scenario. Semistructured interviews were conducted and observations of their interaction with the reviews were recorded (LP).

Participants were asked to 'think aloud' as they completed the task and these comments were audio recorded. Both prototypes were presented to each physician and a random sequence was generated by a computer technician using MySQL's rand() function³¹ so that the order in which the prototypes were presented were randomised.

At the end of this session, participants were asked a series of questions using a semistructured interview guide to learn about user satisfaction and to ask for suggestions on improving the document (see online supplementary appendix A). The System Usability Scale (SUS)^{32 33} was completed by the final eight participants (see online supplementary appendix B). This is a reliable 10-item questionnaire with five response options (from 'strongly agree' to 'strongly disagree') for respondents. This was introduced in the second and third iteration of testing after the tool had been modified. Field notes were also taken during the session (LP).

Content analysis was conducted on the audio-recorded interviews and field notes. After each set of participants, changes were made to the prototypes based on problems identified from the data analysis. The prototypes were modified by a graphic designer. The revised prototypes were then used in the next iteration of usability testing and subsequent refinements were made.

Measures

Case scenarios: The case scenarios were constructed with a human factors engineer based on user tasks such as locating the original source for the document and identifying the number of studies in the systematic review. Although the two prototypes are distinct, essential information appears in the same position for both documents. For example, the table displaying the main results and meta-analysis information is positioned in the same location in the document for the case-based and the evidence-expertise formats. The majority of tasks (13 of 15 case scenarios) given to participants were relevant to prototypes and required users to locate information in the same place for either prototype. A total of two case scenarios were used for only one of the prototypes (one task was unique to the case-based prototype; one task was unique to the evidence-expertise prototype).

An example of a case scenario (relevant to both prototypes) is as follows,

A patient comes in with a clipping from a magazine that suggests tetracycline is the best treatment for more extensive skin lesions on the central part of their face. Can you locate information that would address their question? [Participants were prompted to think aloud and state when they had completed a task].

Since the order in which the two prototypes were presented to participants was randomised and the majority of questions (13 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.

Semistructured interview: The semistructured interview was used to provide further insight into the usability of the two prototypes. The questions were designed to move from general to specific information. The interviewer (LP) used probes to gain more specific information from the participants about the prototypes, for example, "can you give me an example?" The questions asked include the participants' overall impressions of the document, what was liked best, what was liked least and items that should be added.

The SUS is an industry standard that allows the evaluation of a wide variety of products and services, including hardware, software, mobile devices, websites and applications.³³ The wording in this scale is geared towards applications and was altered by replacing the word 'system' with the word 'document'. A score was calculated from the answers given. A SUS score above 68 would be considered above average and below 68 is below average in the evaluation of the usability of a product.³³

Data analysis

The comments from participants during the tasks and the semistructured interviews were transcribed verbatim. Field notes were transcribed and included in the analytic process. Content analysis was done after each iterative cycle. Field notes and transcripts were reviewed and coded independently by two coders (LP, MRK) using a set of codes generated by the interviewer (LP) by initially reviewing terms of usability problems identified by medical information tools.³⁴ A meeting with the second coder was used to identify discrepancies in coding. The number of case scenarios completed was tracked quantitatively.

Rigour and quality

Strategies identified in Lincoln and Guba's framework were used to enhance rigour and quality.³⁵ Probing questions were used during the interviews and prompts were used during the case scenarios to increase the understanding of participants' meaning.³⁶ To increase validity, two investigators analysed and coded the transcripts independently, then discussed discrepancies until agreement was reached. Also, different data sources were used, for example, observations and interviews, plus all procedures were documented in order to create an audit trail.³⁷ This process of triangulation ensured that discoveries and findings emerged from the data through consensus among investigators. Finally, the interviewer had no relationship with any of the participants.

RESULTS

Ten participants took part in three iterative cycles of usability testing. One hundred fifty two recruitment

emails were sent to potential participants. Six physicians were recruited with this method and four through snow-ball sampling, giving a response rate of 7%.

Demographics

The sample (n=10) included five women and five men. Nine participants were aged between 30 and 65 years and one was less than 30 years old (table 1). All participants were family physicians with the majority (90%) having five or more years of experience. Seven of 10 physicians practised in a private clinic. Four participants identified their practice as being in an inner city setting and six indicated that their practice was in an urban/suburban setting.

Usability of the prototypes

Sessions took between 50 and 60 min to complete the case scenarios, interview and SUS. Two participants took part in the first iteration of testing in July 2013 with the original prototypes (see online supplementary appendix C). Recommendations were generated from this analysis and given to the graphic designer who modified the prototypes (see online supplementary appendix D). These recommendations focused on making the information that related to the meta-analysis, the clinical summary and the conditions with no evidence more prominent. These revised prototypes were used in the second iteration of testing which were held in August and September 2013. After three participants finished usability testing, two investigators (LP, MRK) independently coded these transcripts. Minor recommendations were given to the graphic designer and these modifications were implemented. These modifications concentrated on making the explanation for the evidence rating scale distinct and changing the header for the conditions with no trials (see online supplementary appendix E). The modified prototypes were taken into the third iteration of usability testing in January 2014. Five participants took part in usability testing and their transcripts were independently coded by two investigators (LP, MRK) who came to agreement that no new information was being learnt from participants that required major revisions to the prototypes. No further participants were recruited after this set of data was analysed.

The completion of tasks based on the case scenarios were tracked quantitatively (table 2). Changes were made after the first iteration of usability testing based on this information (table 3). Locating the information for the meta-analysis was a challenge for all participants for all three sets of testing. After the first iteration, the graphic designer altered the prototypes in order to make the heading for the meta-analysis more prominent, and when the case scenario for this task was presented to participants in the second iteration of testing, it was emphasised that they were to look for information specific to the meta-analysis. Despite this, during the second iteration of usability testing, participants were not able to locate the

Table 1 Demographics and work profile of study participants

| | Cycle 1–3 participants: N=10 |
|---|---------------------------------|
| Age (years) | |
| <30 | 1 |
| 30–39 | 1 |
| 40–49 | 4 |
| 50–59 | 3 |
| 60–65 | 1 |
| >65 | – |
| Sex | |
| Women | 5 |
| Men | 5 |
| Area of practice | |
| Family medicine | 10 |
| General internal medicine | – |
| Years in practice | |
| <5 | 1 |
| 5–10 | 1 |
| 11–15 | 2 |
| 16–25 | 2 |
| >25 | 4 |
| Work setting* | |
| Private office/clinic (excluding free-standing walk-in clinics) | 7 |
| Community clinic/Community health centre | – |
| Free-standing walk-in clinic | – |
| Academic health sciences centre | 4 |
| Community hospital | – |
| Nursing home/home for aged | – |
| University/faculty of medicine | 1 |
| Administrative office | – |
| Research unit | – |
| Free-standing laboratory/diagnostic clinic | – |
| Other (1 rehabilitation centre; 2 emergency room) | 2 |
| Practice population | |
| Inner city | 4 |
| Urban/suburban | 6 |
| Small town | – |
| Rural | – |
| Geographically isolated/remote | – |
| Other | – |

*Participants listed all settings where they worked (up to three settings per person).

meta-analysis information. Further modifications were made by the graphic designer; however, two of the final five participants in the third iteration of testing did not locate this information. Participants in the first iteration of usability testing could not locate the ‘take-home’ message. Modifications were made to make this information more distinct, and the wording of the case scenario given to participants was altered to clarify that the phrase ‘take-home’ message was a synonym for the word summary. This ceased to be an issue in the second and third iterations of testing and was completed by these participants without

Table 2 Case scenario completion

| Task | Number of tasks not completed | | |
|---|-------------------------------|----------------|----------------|
| | Cycle 1 n=2 | Cycle 2 n=3 | Cycle 3 n=5 |
| Information in same location for both prototypes: | | | |
| Number of studies contributing to meta-analysis <i>Task given to participants:</i> A colleague briefly describes the meta-analysis. Can you locate this information and describe how many studies it summarises? | 2 (100%) | 3 (100%) | 2 (40%) |
| Locate evidence rating scale <i>Task given to participants:</i> You want to present some of this information to others and need to confirm how the evidence rating scale works. Can you describe the evidence rating scale? | | | 1 (20%) |
| Locate adverse event information <i>Task given to participants:</i> You would like to prescribe rilmenidine and want to confirm any adverse events. Can you locate this information and describe it? | | 1 (33%) | |
| State 'take-home message' <i>Task given to participants:</i> You need to give a group of residents the clinical summary or a 'take-home message' from the document. What information would you share with them? | 2 (100%) | | |
| Report explanation for odds ratio <i>Task Given to Participants:</i> You want to confirm the explanation for an odds ratio. Can you find this information and describe it? | | 1 (33%) | |
| Treatment for ocular rosacea <i>Task given to participants:</i> There are no trials for several treatments. Can you locate one treatment for ocular rosacea where there are no randomised controlled trials for evaluation and state it? | | 1 (33%) | 2 (40%) |
| Information in different locations on prototypes: | | | |
| Number of studies contributing to review <i>Task given to participants:</i> You want a sense of how many studies contributed to the entire review. Locate this information and describe the number of studies | 1 (50%) | 2 (66%) | 1 (20%) |
| Total errors | 5 | 8 | 6 |

error. There were two case scenarios that were completed by all participants in the first iteration of testing; however, these same tasks were not completed by at least one participant in the second iteration of testing. One participant was not able to locate information about trials that had no evidence and attention was given to this by the graphic designer between the second and third iterations of testing. Two participants were not able to complete this task during the third iteration of testing; however, no further modifications were made as earlier work with focus groups¹⁵ had indicated that this information had lesser importance which limited graphics in their options for any further modifications, that is, there were no further ways to graphically display the information. Also, one participant could not locate information describing an odds ratio during the second iteration of testing. To address this, the three pieces of information offered along the bottom of the page (information describing the odds ratio, trials with no evidence, and the strength of evidence scale legend) were presented more distinctly as three separate units by using boxes and white space. Despite this, one participant stated that they could not complete the task asking them to locate the evidence rating scale legend

in the third iteration of testing. No changes were made to the prototypes as on review of transcripts it was identified that the question may have been misinterpreted by the participant. The question was worded as, 'Can you describe the evidence rating scale?' (to prompt the participant to describe the scale as strong, moderate and weak). The participant's answer suggests that they may have thought that 'evidence rating scale' was a formal term as they stated, "I don't actually know what that term means... I think it means what kind of statistical analysis was done and I would imagine it has to do with this here, the odds ratio."

The SUS was administered in the second and third iterations of testing (table 4). An SUS score above 68 would be considered above average and anything under 68 is below average.³³ The case-based shortened systematic review received an 'above average' score of 72.5 and 80 (2 of 3 participants) in the second iteration, and 87, 90 and 97.5 (3 of 5 participants) in the third iteration of testing. One participant gave it a score of 52.5 (below average) in the second iteration, and two people scored it 52.5 and 67.5 during the third iteration. All participants scored the evidence-expertise shortened systematic review 'above

Table 3 Changes made to prototypes**Task**

Number of studies contributing to meta-analysis

Task Given to Participants: A colleague briefly describes the meta-analysis. Can you locate this information and describe how many studies it summarizes?

Iteration/ Errors:

Cycle 1=2 errors – Change made: Heading moved from right-hand side to left-hand side

Cycle 2=3 errors – Change made: Text in heading bolded

Cycle 3=2 errors – Change made: *No further changes*

Task

Locate evidence rating scale

Task Given to Participants: You want to present some of this information to others and need to confirm how the evidence rating scale works. Can you describe the evidence rating scale?

Iteration/ Errors:

Cycle 3=1 error – Change made: *No changes made - review of transcripts indicated misinterpretation of question*

Task

State 'take home message'

Task Given to Participants: You need to give a group of residents the clinical 'take-home message' from the document. What information would you share with them?

Iteration/ Errors:

Cycle 1=2 errors – Change made: Question re-worded to 'summary or take-home message'

Task

Report explanation for odds ratio

Task Given to Participants: You want to confirm the explanation for an odds ratio. Can you find this information and describe it?

Iteration/ Errors:

Cycle 2=1 error – Change made: Information on bottom of document distinctly separated by adding boxes or white space

Task

Treatment for ocular rosacea

Task Given to Participants: There are no trials for several treatments. Can you locate one treatment for ocular rosacea where there are no randomized controlled trials for evaluation and state it?

Iteration/ Errors:

Cycle 2=1 error – Changes made: i) Text: colour changed from red to black; capitalized; made larger, and ii) Information on bottom of document distinctly separated by adding boxes or white space

Cycle 3=2 errors – Change made: *No further changes*

Task

Number of studies contributing to review

Task Given to Participants: You want a sense of how many studies contributed to the entire review. Locate this information and describe the number of studies.

Iteration/ Errors:

Cycle 1=1 error – Change made: Efforts focused on making number of studies in the meta-analysis distinct

Cycle 2=2 errors – Change made: Efforts focused on making number of studies in the meta-analysis distinct

Cycle 3=1 error – Change made: *No further changes*

average' with scores of 77.5, 87.5 and 95 (second iteration) and 100, 77.5, 75, 70 and 70 (third iteration).

DISCUSSION

The usability testing revealed features of the shortened systematic review prototypes that prompted modifications. Primary care physicians provided information on how the prototypes could be improved for more effective use. Changes that were made focused on making information prominent so that it could be found more readily. Two items that were consistently challenging to participants were locating the number of studies that

contributed to the meta-analysis, and the number of studies that contributed to the entire systematic review. It is possible that physicians were not clear about the distinction between a systematic review and meta-analysis. Research indicates that physicians have low levels of knowledge with regard to research methodology and data analysis,^{38–41} and this may have contributed to making this case scenario challenging to complete.

The SUS scores were above average in all instances with the exception of three participants and their scoring of the case-based shortened systematic review. The transcripts were reviewed for these participants to identify anything that required further attention. On initially

Table 4 System usability score: cycle 2 and cycle 3 participants

| | Participant number | SUS score |
|------------------------------|--------------------|-----------|
| Cycle 2 | | |
| Evidence-expertise prototype | 3 | 95 |
| | 4 | 77.5 |
| | 5 | 87.5 |
| Case-based prototype | 3 | 52.5 |
| | 4 | 80 |
| | 5 | 72.5 |
| Cycle 3 | | |
| Evidence-expertise prototype | 6 | 100 |
| | 7 | 77.5 |
| | 8 | 70 |
| | 9 | 75 |
| | 10 | 70 |
| Case-based prototype | 6 | 97.5 |
| | 7 | 67.5 |
| | 8 | 90 |
| | 9 | 52.5 |
| | 10 | 87.5 |

Scores have a range of 0–100.
SUS, System Usability Scale.

viewing the case-based prototype, all three participants immediately declared a dislike for the case-based prototype. Examples of this included proactively stating a preference for the evidence-expertise version, repeating this preference more than once in the post-test interview, and stating that they would prefer to have it formatted exactly as the evidence-expertise prototype. Given this clear preference, it was decided not to make further alterations to the prototype based on these low SUS scores.

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 randomised 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 in a clinical scenario.

Study limitations

The development process for the prototypes described in this paper needs to be considered within the context of certain limitations. The sample was small, which could be considered a threat to the generalisability of the results. However, it has been shown that with as few as five participants, the majority of usability problems and issues can be identified with a representative sample of end users.⁴² Also, the demographics of the sample provide some indication of a diverse group in that there is an equal representation of men and women, practising from less than 5 years to more than 25 years. Information on training in critical appraisal or experience in conducting systematic reviews was not collected from participants who may have indicated levels of

expertise related to evidence-based information tools. However, these data would not have a great impact with regard to the usability testing as problems encountered with the use of the tool were being assessed rather than the content or comprehension. Estimates have shown that a single usability testing cycle of design/evaluation/redesign can lead to as much as a 10-fold reduction in usability problems.^{43 44} Moreover, we found that we were not obtaining new information with our interviews in the third iteration of testing and thus had reached saturation. Also, the sample was too small to allow the examination of whether personal (eg, age) or experiential factors were related to the use of the shortened systematic reviews. Future work may help determine how trait characteristics of users impact usability measures.

CONCLUSION

Iterative cycles of multifaceted usability testing provided insight into areas that needed to be refined for two formats (case-based and evidence-expertise) that represent a summary of a full-length systematic review. Usability testing included giving case scenarios to participants that involved them completing tasks of locating information or items within the prototypes, semistructured interviews and, for a portion of participants, completing a systematic usability scale. Changes were made by a graphic designer to the prototypes based on recommendations generated from the analysis of the usability testing. Alterations focused on making information more prominent so that information could be readily located by users. Conducting usability testing during the prototype phase with users provides the opportunity to address usability issues that may impact use of documents or tools. Incorporating their input before finalising the prototypes is done with the aim of increasing their usability and potential use in healthcare.

Contributors SES conceived of the idea. LP conducted the usability testing. LP and MRK performed the coding and data analysis. LP wrote the manuscript and all authors provided editorial advice.

Funding This work was supported by the Canadian Institutes of Health Research.

Competing interests None.

Ethics approval Ethics approval was obtained from the University of Toronto and St. Michael's Hospital, Toronto, Canada Research Ethics Review Boards.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Cochrane Collaboration. What are Cochrane Reviews? <http://www.cochrane.org/about-us/newcomers-guide#reviews> (accessed 7 Apr 2014).

2. Guyatt GH, Sackett DL, Sinclair JC, *et al.* 'Users' guides to the medical literature: IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995;274:1800–4.
3. Laupacis A, Straus S. Systematic reviews: time to address clinical and policy relevance as well as methodological rigor. *Ann Intern Med* 2007;147:273–4.
4. De Vito C, Nobile CG, Furnari G, *et al.* Physicians' knowledge, attitudes and professional use of RCTs and meta-analyses: a cross-sectional survey. *Eur J Public Health* 2009;19:297–302.
5. Ely JW, Osheroff JA, Ebell MH, *et al.* Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ* 2002;324:710.
6. Ely JW, Osheroff JA, Chambliss ML, *et al.* Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Assoc* 2005;293:217–24.
7. Fozi K, Teng CL, Krishnan R, *et al.* A study of clinical questions in primary care. *Med J Malaysia* 2000;55:486–92.
8. D'Alessandro DM, Kreiter CD, Peterson MW. An evaluation of information seeking behaviors of general pediatricians. *Pediatrics* 2004;113:64–9.
9. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc* 2006;94:55–60.
10. Andrews JE, Pearce KA, Ireson C, *et al.* Information-seeking behaviors of practitioners in a primary care practice-based research network (PBRN). *J Med Libr Assoc* 2005;93:206–12.
11. Perrier L, Mrklas K, Shepperd S, *et al.* Interventions encouraging the use of systematic reviews in clinical decision-making: a systematic review. *J Gen Intern Med* 2011;26:419–26.
12. Rosenbaum SE, Glenton C, Nylund HK, *et al.* User testing and stakeholder feedback contributed to the development of understandable and useful summary of findings tables for Cochrane reviews. *J Clin Epidemiol* 2010;63:607–19.
13. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 2010;63:620–6.
14. Perrier L, Persaud N, Ko A, *et al.* Development of two shortened systematic review formats for clinicians. *Implement Sci* 2013;8:68.
15. Perrier L, Kealey MK, Straus SE. An iterative evaluation of two shortened systematic review formats for clinicians: a focus group study. *J Am Med Assoc* 2014;311(12):e341–6.
16. Atkinson RA, Derry SJ, Renkl A, *et al.* Learning from examples: Instructional principles from worked examples research. *Rev Educ Res* 2000;70:181–214.
17. Lee AY, Hutchison L. Improving learning from examples through reflection. *J Exp Psychol Appl* 1998;4:187–210.
18. Robertson I, Kahney H. The use of examples in expository texts: outline of an interpretation theory for text analysis. *Instruc Sci* 1996;24:93–123.
19. Moreno R, Mayer RE. Engaging students in active learning: the case for personalized multi-media messages. *J Educ Psych* 2000;92:724–33.
20. Czuchry M, Dansereau DF. The generation and recall of personally relevant information. *J Exp Psych* 1998;66:293–315.
21. Thomas JC. Story-based mechanisms of tacit knowledge transfer. 2002. ECSCW 2001 Workshop on Managing Tacit Knowledge.
22. Sackett DL, Rosenberg WM, Gray JA, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
23. El Dib RP, Atallah AN, Andriolo RB. Mapping the Cochrane evidence for decision making in health care. *J Eval Clin Pract* 2007;13:689–92.
24. Lottridge DM, Chignell M, Danicic-Mizdrak R, *et al.* Group differences in physician responses to handheld presentation of clinical evidence: a verbal protocol analysis. *BMC Med Inform Decis Mak* 2007;7:22.
25. Health Information Research Unit. McMaster PLUS. http://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx (accessed 15 Feb 2013).
26. van Zuuren EJ, Gupta AK, Gover MD, *et al.* Systematic review of rosacea treatments. *J Am Acad Dermatol* 2007;56:107–15.
27. U.S. Department of Health & Human Services. Glossary. <http://www.usability.gov/what-and-why/glossary/u/index.html> (accessed 7 Apr 2014).
28. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. *Sociol Methods Res* 1981;10:141–63.
29. Snodgrass A, Coyne R. Models, metaphors and the hermeneuticals of designing. *Des Issues* 1992;9:56–74.
30. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. *Proceedings of ACM INTERCHI'93 Conference*; Amsterdam, The Netherlands, 24–29 April 1993:206–13.
31. MySQL. Mathematical Functions. http://dev.mysql.com/doc/refman/5.0/en/mathematical-functions.html#function_rand (accessed 7 Apr 2014).
32. Brooke J. SUS—a quick and dirty usability scale. 1986. hell.meiert.org/core/pdf/sus.pdf (accessed 7 Feb 2014).
33. System Usability Scale. <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (accessed 7 Apr 2014).
34. Kushniruk AW, Triola MM, Borycki EM, *et al.* Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. *Int J Med Inform* 2005;74:519–26.
35. Lincoln Y, Guba EG. *Naturalistic inquiry*. Newbury Park, CA: Sage Publications, 1985.
36. Nelson AM. Addressing the threat of evidence-based practice to qualitative inquiry through increasing attention to quality: a discussion paper. *Int J Nurs Stud* 2008;45:316–22.
37. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ* 2000;320:50–2.
38. Wegwarth O, Schwartz LM, Woloshin S, *et al.* Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012;156:340–9.
39. Baldwin LM, Trivers KF, Matthews B, *et al.* Vignette-based study of ovarian cancer screening: do U.S. physicians report adhering to evidence-based recommendations? *Ann Intern Med* 2012;156:182–94.
40. Novack L, Jotkowitz A, Knyazer B, *et al.* Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors. *Postgrad Med J* 2006;82:817–22.
41. Raina PS, Brehaut JC, Platt RW, *et al.* The influence of display and statistical factors on the interpretation of metaanalysis results by physicians. *Med Care* 2005;43:1242–9.
42. Nielsen J. How to conduct a heuristic evaluation. <http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation> (accessed 7 Apr 2014).
43. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. *BMJ* 2004;328:1150.
44. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Comput Biol Med* 2002;32:141–9.