How does Evidence Affect Clinical Decision-making?

Paul Fontelo, MD, MPH,

National Library of Medicine, 8600 Rockville Pike, Bethesda, Maryland 20894

Fang Liu, MS, and

National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894

Raymonde C. Uy, MD, MBA

National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894

Abstract

In 1998, the "Evidence Cart" was introduced to provide decision-support tools at the point of care. A recent study showed that a majority of doctors who previously stated that evidence was not needed sought it nevertheless when it was easily available. In this study, invited clinicians were asked to rate the usefulness of evidence provided as abstracts and "the bottom-line summaries" (TBL) using a modified version of a Web app for searching PubMed and then specify reasons how it might affect their clinical decision-making. The responses were captured in the server's log. One hundred and one reviews were submitted with 22 reviews for abstracts and 79 for TBLs. The overall usefulness Likert score (1=least useful, 7=most useful) was 5.02 ± 1.96 (4.77 ± 2.11 for abstracts and 5.09 ± 1.92 for TBL). The basis for scores was specified in only about half (53/101) of reviews. The most frequent single reason (32%) was that it led to a new skill, diagnostic test, or treatment plan. Two or more reasons were given in 16 responses (30.2%). Two-thirds more responders used TBL summaries than abstracts confirming further that clinicians prefer convenient easy-to-read evidence at the point of care. This study seems to show similar results as the Evidence Cart study on the usefulness of evidence in clinical decision-making.

Keywords

evidence-based medicine; clinical decision-making; Web app; journal abstract; "the bottom line"

INTRODUCTION

In 1998, Sackett and Strauss used an "Evidence Cart" containing MEDLINE, Cochrane Library, Critically Appraised Topics (CATs), and other references on compact disks and a notebook computer to provide evidence during rounds.[1] They found that in the majority (81%) of instances when the cart was used, the evidence sought could affect how a patient might be managed (diagnosis and treatment). In 71 successful searches which were

"assessed from the perspective of most junior team members responsible for each patient's evaluation and management", 52% of the information found "confirmed their current or tentative diagnostic or treatment plans" while 25% "led to a new diagnostic skill, an additional test, or a new management decision" and 23% "corrected a previous clinical skill, diagnostic test, or treatment". When the cart was not easily available, there was still a high level of perception of need but searching for evidence was only done in 12% of instances. This is not unusual because the greater the effort is required to find evidence, the less likely it will be used.[2–4] If evidence is not easily accessible, clinicians will not seek for it even if it's needed. They concluded that, "Making evidence quickly available to clinicians on a busy medical inpatient service using an evidence cart increased the extent to which evidence was sought and incorporated into patient care decisions".[1]

In a recent study among resident physicians in a tertiary care hospital in a developing country, the majority of physicians who initially said that evidence was not needed (17%) sought it nonetheless when made available (83%).[5] Moreover, it showed that journal abstracts and full text articles equally increased the accuracy of clinical decisions in clinical simulations. Internal medicine residents reported that information found online affected clinical decision in 78% impacting their confidence, knowledge and care of patients in the future.[6]

A limitation of the evidence cart was that it could not be carried to bedside rounds because it was too bulky. Today, the evidence cart has been replaced by wireless mobile devices that can be conveniently carried everywhere and access information anywhere in the world. However, this technology brings a different type of "bulk" – the massive volume of information that search engines can discover on the Internet. Although the evidence cart in Sackett and Strauss' report found useful information within 10.2 – 25.4 seconds, the challenge today is finding high quality, reliable and useful information within the deluge of resources that mobile devices can discover in a third or less of the time it took to find it in the evidence cart. This highlights the importance for physicians to have good information seeking skills, such as proper keywords, methods, and reliable Web sites.

Many Evidence-Based Medicine (EBM) Websites provide links to reliable evidence summaries and pre-appraised resources on the Internet. Windish recently published a list of the apps that are useful for answering clinical questions.[7] Although Electronic Medical Records (EMR) systems deployed in hospitals and clinics today also have integrated clinical decision support tools, a recent meta-analysis of 28 randomized controlled trials did not seem to indicate an effect on mortality overall and only afforded small reductions in cost-savings and minor increases in health services utilization in economic outcomes studies.[8] A significant effect was shown in preventing morbidity but selective outcome reporting and publication bias could not be ruled out.

Points of care decision support tools provide clinicians with useful information in informing about patient care. It works both ways and has also been shown to improve communication between healthcare provider and patient.[9] Many free and subscriber-based resources, such as, AHRQ's National Guideline Clearinghouse, Cochrane Library, UptoDate, and DynaMed are available on the Internet and as downloadable apps for smartphones and other mobile

devices. However, because of time constraints in clinical practice, information overload, access restrictions to resources, and lack of evidence appraisal skills, clinicians may still be unable to practice evidence-based medicine.[10]

A study by Smith published in BMJ in 1996 reviewed 13 studies involving more than 2300 participants consisting mostly of American physicians on the information needs and wants of doctors.[4] One of the summary points states that, "New information tools are needed: they are likely to be electronic, portable, fast, easy to use, connected to both a large valid database of medical knowledge and the patient record, and a servant of patients as well as doctors". The usefulness of medical information is related to its relevance, validity, and the amount of work to acquire it.[2–4] The more difficult it is to find, the less likely it will be sought as demonstrated in the Evidence Cart study. The need for evidence at the point of care has not changed today as shown by the proliferation of point of care tools, apps for mobile devices like smartphones and tablets of summaries, synthesized and digested medical information.[11]

AIMS

In this study, we wanted to find out how evidence provided as abstracts and TBLs ("the bottom line" summaries)[10] and accessed with mobile devices in wireless networks affects clinical decision-making.

METHODS

The "PubMed for Handhelds" (PubMed4Hh) index page (http://pubmedhh.nlm.nih.gov) was modified (http://pubmedhh.nlm.nih.gov/indexpar.html) so clinicians could rate the usefulness of abstracts and TBLs ("the bottom line") summaries. The Web interface was designed just like the regular PubMed4Hh Website whereby users could choose search tools like PICO (Patient/Problem, Intervention, Comparison, and Outcome), askMEDLINE, Consensus Abstracts and Journal Browser to search for PubMed citations.

Invitations to participate in the study were circulated among online social media groups and instant messaging (IM) groups composed of physicians, which included general practitioners, residents in training, and consultants. Direct invitations were also sent to some physicians and training program directors using either e-mail or instant messaging applications. Since most of these invitees were regular users of PubMed4Hh, they were directed to the modified index page (see above) and search as they normally would. Functionality of the modified search was the same. The only difference is when they read the search results, a rating scale (Figure 1) for grading their usefulness would appear below the abstract or TBL.

Each citation in the results page showed both the TBL and abstract. A 7-point Likert scale (1=Least useful to 7=Most useful) form was provided for users to rate the usefulness of the TBL or the abstract (Figure 1). After the score was submitted, a feedback form (Figure 2) was shown, patterned after the Evidence Cart study[1] to indicate the rationale for their scores. The type of questions asked were sorted to foreground and background questions according to the guides available on the Internet.[14, 15] The search mode used (PICO or

askMEDLINE) and summary read (abstract or TBL) score and the ratings data were stored in a database anonymously.

The mode and frequency of Likert scores (ordinal data) for TBL and abstract were calculated for each of the questions asked. In this study, we also assumed the space intervals between adjacent scores were symmetrical so interval levels of measurements including mean and standard deviation were also performed.[16, 17]

The NIH Office of Human Subjects Research had designated this research as exempt from IRB review.

RESULTS

Server logs were reviewed during the study period between June 20, 2013 to February 28, 2015. Although the total time period spans 20 months, data collection only cover a total of 12.5 months in three distinct time periods: Period 1: 2013-06-20 to 2013-10-11; Period 2: 2013-02-05 to 2013-03-07; Period 3: 2014-06-26 to 2015-02-28. The total number of searches and responses for the three periods were 31, 10 and 60, respectively. Reminders and invitations to participate in the study were sent when the searches and reviews waned, hence the gaps in data collection.

One hundred one entries were logged through the modified index page. These entries came from 12 countries with the majority coming from the United States (65). Other countries represented were: the Philippines, 10; Singapore, 9; Mexico, 4; Brazil and Hong Kong, 3 each; Puerto Rico, 2; and one each for Australia, Canada, Indonesia, Qatar, and the Netherlands. In comparison, 113293 searches were logged on the regular PubMed4Hh server during the same 20-month period.

Some study participants reviewed one or more TBLs or abstracts for some questions thus the number of reviews exceeded the number of searches (101 vs. 63). For example, a participant searching a clinical question such as the comparison of the effect of lifestyle change and medication for the prevention of type 2 diabetes mellitus may retrieve a 100 or more citations. Each of these citations may have a TBL and an abstract. If they rated both the TBL and abstract of a particular journal citation, they would have accounted for 2 reviews although only one question was searched.

These queries included 51 foreground or patient specific questions and 12 background or general knowledge questions. The majority of searches (56) were questions on therapy while seven were related to diagnosis. There were no searches involving etiology or prognosis.

Figure 3 summarizes the usefulness ratings of study participants who reviewed 22 abstracts and 79 TBLs. The Likert scale mode for abstracts and TBLs were both "7" (6/22 or 27.27% for abstracts and 27/79 or 34.18% for TBLs).

When calculated using interval estimates, the means for abstracts and TBLs were 4.77 and 5.09 with 95% CIs [3.89, 5.65] and [4.67 to 5.51] respectively. Overall, the mean for both abstracts and TBLs was 5.02 with CIs [4.64 to 5.40].

Seven of submitted responses were searches done using non-English terms (five in Spanish and one in Portuguese) with mean Likert scores of 1.57 and 95% CIs 1.17 to 1.97. These scores (all rated TBLs) were excluded.

In only 53 of 101 of usefulness reviews were rationales provided for the Likert scale ratings (Table 1). The six who selected "none of the above" viewed an equal number of abstracts and TBLs with Likert scale scores of "1", "2", "4" and "3", "5", "7", respectively.

Combinations of reason for Likert scale scores were also measured (Table 2). When "confirmed my current or tentative diagnostic or treatment plan", and "corrected my previous clinical skill, diagnostic test, or treatment plan" were combined with one or two other reasons and "All of the above", they accounted for 22 and 20 reviews, respectively. "Led to new diagnostic skill, additional test, new management decision" combined with others was the most predominant reason mentioned in 33 reviews.

Table 3 shows an evaluation of evidence scores based on the number of citations reviewed. Eighty-eight percent (89/101) of scores were based on the review of only one citation.

DISCUSSION

Although it was only original envisioned as a device of indexing journal articles, the abstract has now become for many, the only section of the journal article read and maybe even for some, used for informing clinical decisions.[18–21] This is especially prevalent in low-resource regions of the world where full text articles may not be available whereas abstracts are. Smartphones and the ubiquitous mobile Internet have contributed further to its popularity. Comparisons made between full-text articles and abstracts have shown that they may contain inaccuracies but these were minor errors and did not seem to affect the clinical bottom line.[18] Search tools, such as Consensus Abstracts have been developed to present abstracts and TBLs assembled and viewed together for convenient reading and also to improve their validity.[10]

Although the modified PubMed4Hh interface allows users to view TBLs, abstracts and provides links back to the full-text articles in PubMed Central and publishers' sites, we are only able to compare the participants' scores and feedback for TBLs and abstracts because tracking full-text preferences is not possible. However, through this approach, we aimed to determine further the usefulness of abstracts and TBLs for informing clinical decision-making and the participants' preferences between the two summaries.

In this study, we wanted to find out how the use of evidence affects decision-making and whether it has changed since the 1997 Evidence Cart study done before the use of smartphones and mobile networks.[1] In the majority of instances, evidence affects clinical decisions in many ways, as this study confirms. Individually, the choices "confirmed my current or tentative diagnostic or treatment plan", "led to new diagnostic skill, diagnostic test, or treatment plan only" and "corrected my previous clinical skill, diagnostic test, or treatment plan" accounted only for 9, 17, and 5 reviews, respectively. Unlike the original study, the current study allowed participants to select more than one, all, or none of the choices. Consequently, when considered with other options (Table 2), they increased to 22,

33, and 20 reviews, respectively. "Led to new diagnostic skill, diagnostic test, or treatment plan only" was the most predominant reason accounting for only for 17 reviews individually but increased to 33 when combined with other choices.

Although no direct comparison can be made with the Evidence Cart study because of the greater number of choices of rationale for grading evidence in this study, when the same options were compared, "led to new diagnostic skill, diagnostic test, or treatment plan only" accounted for the majority of selections (32.1% vs. 25% in Sackett study), followed by "confirmed my current or tentative diagnostic or treatment plan" (17% vs. 52%) and "corrected my previous clinical skill, diagnostic test, or treatment plan" (9.4% vs. 23%). We are uncertain how to explain the shift, but perhaps one possible reason is that information accessed through the Internet today may be providing more diagnostic test and treatment management options since the Sackett and Strauss study published 17 years prior. Alternatively, perhaps, this is simply a finding that may not be validated in a subsequent study. It might be research topic worth pursuing whether there are really more tests and treatment options.

TBL summaries were greatly preferred over abstracts by almost 4:1. An explanation may be found in a study, "What clinical information do doctors need?" where Smith observed that, "the best information sources provide relevant, valid material that can be accessed quickly and with minimal effort".[4] TBLs provide easily readable information that captures the journal article's essential information.[19–21] This is also supported by feedback from app users who have said that the TBL is the "best" feature of PubMed4Hh app because it delivers easy-to-read essential information quickly.

Since this was an unobserved study, we are unable to determine whether these searches were actual clinical questions that arose during rounds or any other clinical activity. However, the observation that the majority of searches (56) were questions on therapy while seven were related to diagnosis, may suggest that these were questions might be related to clinical activity. The questions included 51 foreground questions and 12 background questions suggesting that the participants were more aware of EBM practice and principles and the formulation of well-built clinical questions.[22]

This study also suggests that clinicians' preference for relevant, valid material that can be accessed quickly and conveniently.[4] The TBL summary provides a concise, easy-to-read summary that captures the essential information of the journal article. TBLs are brief summaries that convey the essential elements of the journal abstract.[23] Around 12% of PubMed's 24 million citations have abstracts. Briefly, it is generated as follows: if the journal abstract follows the IMRAD (Introduction, Methods, Results and Discussion) format and contains the PubMed tag 'Conclusion' or the word 'conclusion', the conclusion or the segment or sentence with "conclusion" is the TBL. In unstructured abstracts, a TBL algorithm developed by the authors parses the sentences, counts the occurrence frequency of the words, and then selects the top five most repeated words as "key words". The sentence with the most number of key words and the last two sentences of the abstract is 'the bottom line'. [24]

The PICO format makes it convenient to formulate searches that will find current therapy and diagnostic workup relevant to patient care. There seems to be a suggestion also, that there is a shift in how evidence affects decision-making from confirming current or tentative practice to leading clinicians to newer patient management skills. The generalizability of these observations needs to be validated by real-time clinical studies.

LIMITATIONS

We believe that the majority of queries and responses are from clinicians since the invitations were directed to physician groups, but we cannot be certain that all responses are. Although we can track whether the participants "clicked" and rated the TBL or Abstracts, full-text links go to the journal publishers' sites or to PubMed Central, which is totally outside of our control. Another limitation is that the participants in this study may be more technological savvy and therefore not representative of the general medical community. Hence, the results of this study may not be universally generalizable and should be interpreted with caution. Also added to the limitations of this study is that we cannot be certain that the queries and responses captured were done by clinicians or physicians at all since this was an unobserved study on the Internet.

CONCLUSION

Clinicians prefer brief summaries of evidence or bottom-line statements accessible at the point of care. It appears also from this study that clinicians regard evidence that leads to new diagnostic skill, diagnostic test, or treatment plan as the primary reason for the usefulness of evidence although a combination of reasons is also high on the list. Mobile or Web applications such as PubMed4Hh may be useful for providing healthcare professionals with a quick way to access online medical literature in support of the practice of evidence-based medicine. Through the app, abstracts and TBLs accessed via a PICO search and askMEDLINE may have an important role in integrating current medical evidence into the clinical practice workflow.

Acknowledgments

References

1. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA. 1998 Oct 21; 280(15):1336–8. [PubMed: 9794314]

2. Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. J Fam Pract. 1994; 39:489–99. [PubMed: 7964548]

3. Slawson DC, Shaughnessy AF. Obtaining useful information from expert based sources. BMJ. 1997; 314:947–9. [PubMed: 9099121]

4. Smith R. What clinical information do doctors need? BMJ. 1996; 313:1062–8. [PubMed: 8898602]

5. Marcelo A, Gavino A, Isip-Tan IT, et al. A comparison of the accuracy of clinical decisions based on full-text articles and on journal abstracts alone: a study among residents in a tertiary care hospital. Evid Based Med. 2013; 18:48–53. [PubMed: 22782923]

6. Schilling LM, Steiner JF, Lundahl K, Anderson RJ. Residents' patient-specific clinical questions: opportunities for evidence-based learning. Acad Med. 2005 Jan; 80(1):51–6. [PubMed: 15618093]

7. Windish D. EBM apps that help you search for answers to your clinical questions. Evid Based Med. 2014 Jun; 19(3):85–7. [PubMed: 24530997]

8. Moja L, Kwag KH, Lytras T, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health. 2014; 104(12):e12–22.

9. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, Llewellyn-Thomas H, Lyddiatt A, Légaré F, Thomson R. Decision aids for people facing health treatment or screening decisions. Cochrane Database of Systematic Reviews. 2011; (10) Art. No.: CD001431. doi: 10.1002/14651858.CD001431.pub3

10. Fontelo P. Consensus abstracts for evidence-based medicine. Evid Based Med. 2011 Apr; 16(2): 36–8. [PubMed: 21427051]

11. Wolters Kluwer Health. [Date accessed 3 August 2015] Physician Outlook Survey. 2013. http://www.wolterskluwerhealth.com/News/Documents/White%20Papers/Wolters%20Kluwer%20Health%20Physician%20Study%20Executive%20Summary.pdf

12. PubMed for Handhelds index page: http://pubmedhh.nlm.nih.gov.

13. PubMed for Handhelds modified index page: http://pubmedhh.nlm.nih.gov/indexpar.html.

14. Weinfeld JM, Finkelstein K, Manag Fam Pract. How to answer your clinical questions more efficiently. 12(7):37–41.

15. Posley, K., Maggio, L. [Date accessed 3 August 2015] Formulating clinical questions. 2015. https://lane.stanford.edu/classes-consult/info-lit-files/PicoE4c.pdf

16. Jamieson S. Likert scales: how to (ab)use them. Med Educ. 2004 Dec; 38(12):1217–8. [PubMed: 15566531]

17. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract. 2010 Dec; 15(5):625–32. [PubMed: 20146096]

18. Fontelo P, Gavino A, Sarmiento RF. Comparing data accuracy between structured abstracts and full-text journal articles: implications in their use for informing clinical decisions. Evid Based Med. 2013 Dec; 18(6):207–11. [PubMed: 23786759]

19. Winker MA. The need for concrete improvement in abstract quality. JAMA. 1999; 281:1129–30. [PubMed: 10188667]

20. Read MEDLINE abstracts with a pinch of salt. Lancet. 2006; 368:1394.

21. Barry HC, Ebell MH, Shaughnessy AF, et al. Family physicians' use of medical abstracts to guide decision making: style or substance? J Am Board Fam Pract. 2001; 14:437–42. [PubMed: 11757886]

22. Richardson, W Scott, MD, Wilson, Mark C., MD MPH, Nishikawa, Jim, MD, Hayward, Robert SA, MD, MPH. The well-built clinical question: a key to evidence-based decisions. ACP Journal Club. Nov-Dec;1995 v123:A12.

23. Tom O, Fontelo P, Liu F. Do computer-generated summaries, 'The Bottom Line (TBL)' accurately reflect published journal abstracts? AMIA Annu Symp Proc. 2007:1135. [PubMed: 18694232]

24. Fontelo P, Liu F, Muin M, Tolentino H, Ackerman M. Txt2MEDLINE: text-messaging access to MEDLINE/PubMed. AMIA Annu Symp Proc. 2006:259–63. [PubMed: 17238343]

Figure 1.
The left panel shows an example of a PICO search. A 7-point Likert scale for rating the TBL or abstract is shown on the right panel.

I rated this TBL/Abstract/Fulltext based on one or all of the following: (Choose one or more)

☐ It confirmed my current or tentative diagnostic or treatment plan

☐ It led to new diagnostic skill, additional test, new management decision

☑ It corrected my previous clinical skill, diagnostic test, or treatment plan

☐ All of the above

☐ None of the above

Submit

Figure 2.
Feedback form patterned after the Evidence Cart study to indicate the rationale for scoring an abstract or TBL.
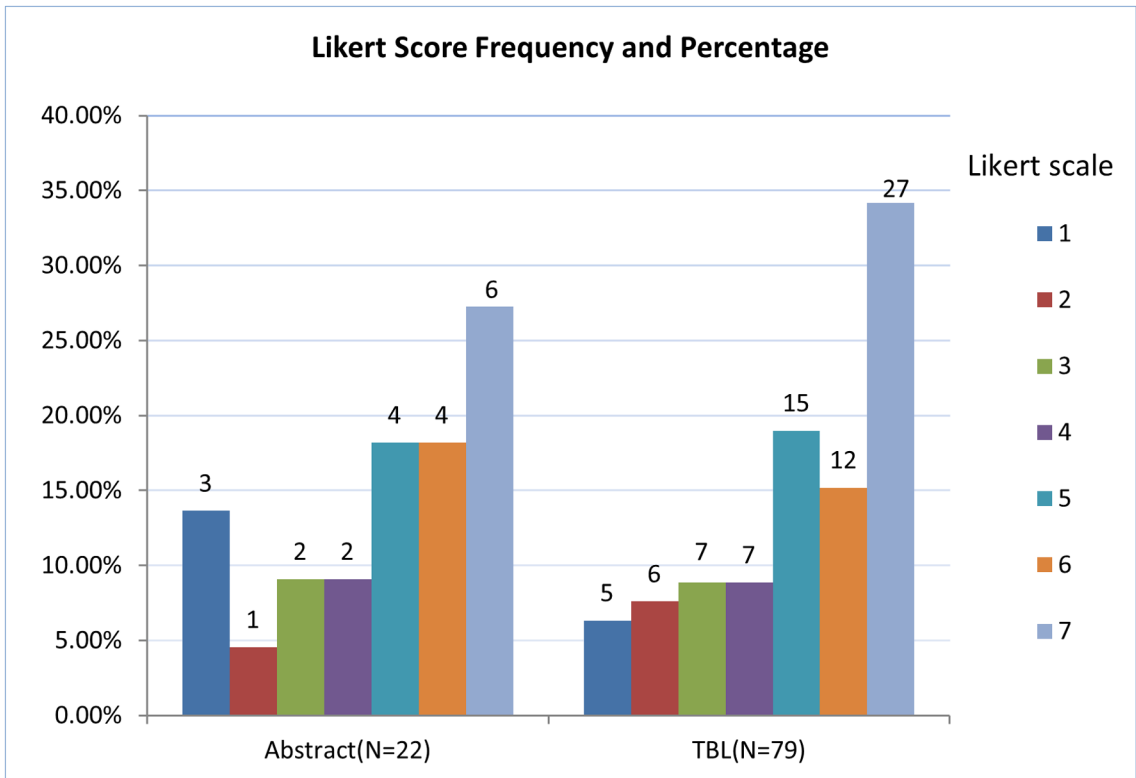
Figure 3.
Likert scale frequencies and percentages of abstracts and TBLs.

Table 1

Count and percentages of rationale for Likert scale scores

Rationale	Count	Percentage of total
Confirmed current or tentative diagnostic or treatment plan only	9	16.98%
Led to new diagnostic skill, diagnostic test, or treatment plan only	17	32.08%
Corrected previous clinical skill, diagnostic test, or treatment plan or management only	5	9.43%
Any two of the above only	4	7.55%
All of the above (Confirmed, Led, Corrected)	12	22.64%
None of the above	6	11.32%
Total	53	100%

Table 2

Percentages of rationale for Likert scale scores

One or a combination of rationale	Count	Percentage of total
Confirmed current or tentative diagnostic or treatment plan plus other reasons	22	29.3%
Led to new diagnostic skill, diagnostic test, or treatment plan plus other reasons	33	44%
Corrected previous clinical skill, diagnostic test, or treatment plan or management plus other reasons	20	26.7%

Table 3

Analysis of the number of citations reviewed in rating evidence.

Number of citations	Number of reviews	Mean, (95% CI)
1	89	5.02, (4.63 to 5.41)
2 to 4	3	6.00, (4.04 to 7.96)
5	6	3.50, (1.31 to 5.69)
More than 5	3	7.00
Total	101	