

Published in final edited form as:

J Clin Epidemiol. 2012 June ; 65(6): 660–668. doi:10.1016/j.jclinepi.2011.12.004.

Two methods provide similar signals for the need to update systematic reviews

Mei Chung^{a,*}, Sydne J. Newberry^b, Mohammed T. Ansari^d, Winifred W Yu^a, Helen Wu^b, Jounghee Lee^a, Marika Suttorp^b, James M. Gaylor^a, Aneesha Motala^b, David Moher^d, Ethan M. Balk^a, and Paul G. Shekelle^{b,c}

^aTufts Evidence-based Practice Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Box 63, 800 Washington Street, Boston, MA 02111, USA

^bThe RAND Health, Southern California Evidence-based Practice Center, Santa Monica, CA

^cWest Los Angeles VA Medical Center, Los Angeles, CA

^dUniversity of Ottawa Evidence-based Practice Center, The Ottawa Methods Centre, Clinical Epidemiology Program, The Ottawa Hospital Research Institute, Ottawa, Canada

Abstract

Objective—Apply and compare two methods that identify signals for the need to update systematic reviews, using three Evidence-based Practice Center reports on omega-3 fatty acids as test cases.

Study Design and Setting—We applied the RAND method, which uses domain (subject matter) expert guidance, and a modified Ottawa method, which uses quantitative and qualitative signals. For both methods, we conducted focused electronic literature searches of recent studies using the key terms from the original reports. We assessed the agreement between the methods and qualitatively assessed the merits of each system.

Results—Agreement between the two methods was “substantial” or better ($\kappa > 0.62$) in three of the four systematic reviews. Overall agreement between the methods was “substantial” ($\kappa = 0.64$, 95% confidence interval [CI] 0.45–0.83).

Conclusion—The RAND and modified Ottawa methods appear to provide similar signals for the possible need to update systematic reviews in this pilot study. Future evaluation with a broader range of clinical topics and eventual comparisons between signals to update reports and the results of full evidence review updates will be needed. We propose a hybrid approach combining the best

© 2012 Elsevier Inc. All rights reserved.

*Corresponding author. Tel.: +617-636-5989; fax: +617-636-8628. mchung1@tuftsmedicalcenter.org (M. Chung).

Publisher's Disclaimer: Disclaimer: The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality (AHRQ) or the U.S. Department of Health and Human Services.

Appendix

Supplementary material

Supplementary material can be found, in the online version, at [10.1016/j.jclinepi.2011.12.004](https://doi.org/10.1016/j.jclinepi.2011.12.004).

features of both methods, which should allow efficient review and assessment of the need to update.

Keywords

Evidence-based methodology; Systematic reviews; Comparative effectiveness reviews; Omega-3 fatty acids; Cardiovascular disease risk factors; Cancer; Cognitive function

1. Introduction

The question of how to determine when a systematic review needs to be updated is of considerable importance. Changes in the evidence can have significant implications for health care practitioners, policy makers, and patients and their caregivers who depend on up-to-date systematic reviews as their foundation. The rapidity with which new research findings accumulate makes it imperative that the evidence be assessed periodically to determine the need for an update. Identifying updating signals would be particularly useful to inform stakeholders where new evidence is sufficient to consider updates of comparative effectiveness reviews (CERs) [1].

Currently, the most commonly used approach to initiating updates is a preset time-based updating frequency [2]. For example, since 2002, the Cochrane Collaboration's policy has been to update Cochrane reviews every 2 years [3]. Such updates involve an investment of time and effort that may not be appropriate for all topics. In 2005, 254 Cochrane updates performed in 2002 were compared with their original reviews published in 1998 [4]. Only 23 (9%) had a change in conclusion, which supports use of a priority approach, rather than an automatic time-based approach, to determine the need for an update [4].

The science of identifying signals for updating systematic reviews has been developing for the past decade. Before 2001, no explicit methods or criteria existed to determine whether evidence-based products remained valid or whether the evidence underlying them had been superseded by newer work. Since the late 1990s, the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) program has been conducting studies to develop methods to assess the need for updating evidence reviews. Two methods have been developed. First, the Southern California EPC (SCEPC) based at the RAND Corporation conducted a study to determine whether AHRQ's clinical practice guidelines needed to be updated and how quickly guidelines go out of date. Based on a conceptual model, the SCEPC developed a method that combines external domain (subject matter) expert opinion with an abbreviated search of the literature published since the original systematic review [5,6]. In 2008, the SCEPC adapted its method to assess the need for updating the CERs that had been prepared to that point (hereafter referred to as "the RAND method") [7]. In parallel, a second method was devised at the University of Ottawa EPC (UOEPC) through assessment of the predictors of the need to update systematic reviews with meta-analyses [8], and the method was then tested using 100 meta-analyses published from 1995 to 2005 [9]. The method did not involve external expert judgment but instead relied on capturing a combination of quantitative and qualitative signals (hereafter referred to as "the Ottawa method").

AHRQ has a rapidly expanding program of comparative effectiveness research, with 195 CERs or evidence reports completed through 2010 and an anticipated 50 CERs to be completed by 2011. Keeping these CERs up to date is a pressing concern for AHRQ and their end users. An efficient method is needed to rapidly and accurately assess the need to update. To this end, AHRQ commissioned the present pilot study to compare the results of the RAND and Ottawa methods for identifying signals for the need for updating. The results of this study would inform the choice of the method for surveillance of signals for updating CERs. Chosen as test cases were three evidence reports on omega-3 fatty acids (omega-3 FAs) conducted by two of the three EPCs conducting this study: the effects of omega-3 FAs for preventing and treating neurological disorders [10], the effects of omega-3 FAs for preventing and treating cancer [11], and the effects of omega-3 FAs on risk factors and intermediate markers for cardiovascular disease (CVD) [12]. The full reports submitted to AHRQ are available online at <http://www.effectivehealthcare.ahrq.gov/>.

2. Methods

In 2004 and 2005, the SCEPC, UOEPC, and the Tufts Medical Center EPC (Tufts EPC) produced nine evidence reports for the Office of Dietary Supplements on omega-3 FAs effects in relation to 17 clinical topics [13]. For the current analysis, three of the original nine evidence reports were chosen because they were of interest to the sponsors of the original project, who wanted information about the possible need to update these reviews while also contributing to the science of determining signals for updating. Thus, two EPCs (SCEPC and Tufts EPC) applied both the RAND and the modified Ottawa methods to assess signals for updating their original reports. This approach was taken to represent a likely future scenario where the group producing the original report would be responsible for assessing the need for updating.

2.1. RAND method

The RAND method uses a four-category scheme (“definitely out of date,” “probably out of date,” “possibly out of date,” or “still valid”) to assess the conclusions and the possible need for updating, based on a combination of external domain expert opinion and an abbreviated search [7]. Table 1 describes the operational definitions of this classification scheme.

For each evidence report (clinical topic), we contacted external domain experts identified from the original reports’ Technical Expert Panels and peer reviewers and other experts known or recommended to us. We aimed to include at least four experts for each report. Separately for each report, we created forms that included the original report’s key questions and conclusions, based on information in the executive summary, discussion chapter, or other summaries (see supplementary Table 1 on the journal’s web site at www.jclinepi.com). We then asked the experts whether, to their knowledge, each of the conclusions was “almost certainly still supported by the evidence.” If the answer was “no,” they were asked to provide any new supporting evidence known to them (see supplementary Tables 2–4 on the journal’s web site at www.jclinepi.com).

For each report, we performed a focused literature search (as described below) and extracted a limited set of data from eligible studies (study design, interventions, sample size, and

relevant study results). We then compared the original report conclusions, a summary of the findings of the new studies, and a summary of the experts' votes and opinions, and from these derived a conclusion about the need to update based on the four-category scheme (Table 1).

2.2. Modified Ottawa method

The Ottawa method relies on a literature search alone from which it arrives at one of three types of signals for the need to update—a qualitative signal, a quantitative signal, or an “other” signal (Table 2)—depending on the body of literature [8]. For the purpose of this study, if the response to a key question in the original report included a meta-analysis, a quantitative signal is sought. If no previous meta-analyses could be conducted, a qualitative or other signal is sought. We did not modify the approach for Ottawa quantitative signal (see supplementary Fig. 1 on the journal's web site at www.jclinepi.com).

For key questions with meta-analyses in the original reports, we first sorted the new studies by sample size. For each new original study, beginning with the largest, we conducted a fixed-effects analysis, pooling the effect size reported in the original meta-analysis and the new study. This process was repeated with each subsequently smaller trial until we found a signal (criteria B1 or B2 in Table 2) or until all new studies were added. The Ottawa method does not aim to recreate the original meta-analysis (which would require collecting data from all meta-analyzed studies). Rather, the original meta-analysis result is entered into a new fixed-effects model as one point estimate. Since the goal of the updated meta-analysis is to find a possible signal for a full update and not to provide a best estimate of the effect, this approach is sufficient.

For key questions without meta-analyses, we searched for “pivotal” trials, defined as trials published in a pivotal journal (see literature search below) or trials published in nonpivotal journals but with at least three times the number of participants as the largest trial in the original report. The findings of such pivotal trials were compared with those of the original reports to determine whether any new findings might invalidate the previous findings or suggest major changes in evidence. If there were no pivotal trials, all relevant new studies were reviewed. For key questions for which there were no prior meta-analyses or randomized controlled trials (RCTs) in the original report, an “other” signal was sought, which was determined based on whether there was a major increase in the number of new studies or at least one new study with at least three times the number of participants as were included in nontrial studies in the original review (Table 2). Further details are available at www.ohri.ca/UpdatingSystRevs.

The original Ottawa method was originally developed and tested on systematic review that had meta-analysis. Because most AHRQ CER key questions do not include meta-analysis, we need to modify the original Ottawa method to deal with this scenario. In collaboration with the original Ottawa method authors (M.T.A., D.M.), we made two modifications to the original Ottawa method for qualitative and other signals: (1) we extended the qualitative signals criteria to include nontrial data for those key questions where the original report included nontrial evidence and (2) we designated new evidence as an A4 signal for updating in a situation where the original report had no evidence.

Regarding the first modification, much of the evidence for cognitive and neurological disease and the entire body of evidence for cancer prevention were from population-based, observational studies. In addition, few of these studies were published in a major general medical journal; thus, it was not possible to meet Ottawa's original definition of a pivotal trial in that respect. In addition, unlike clinical trials where the original study population might be modest in number, a threefold increase in sample size may be unnecessarily difficult to achieve for reviews of large cohort studies. New large cohort studies might be of sufficient interest, regardless of their relative size.

Regarding the second modification, the original Ottawa method is silent on how to assign a signal when the original report had no trials addressing a specific question, and there are new small trials. For this analysis, we designated these situations as A4 (important changes in effectiveness short of "opposing findings"), reasoning that had the Ottawa method been required to consider this situation, the method would have judged the existence of new evidence as being a signal for updating.

2.3. Literature search

Both the RAND and the modified Ottawa methods began with the same literature search. All literature searches were conducted by the UOEPC using the same search terms used for the original reports [10–12]. Because the goal of this exercise was to find signals, not to fully update previous systematic reviews, the searches were limited to five major (pivotal) general-interest medical journals (*Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *the Lancet*, and *the New England Journal of Medicine*, as defined by the original RAND methodology [6]) supplemented, separately for each topic, with a small number of specialty journals that were most frequently cited in the original reviews. Because there is a gap in publication date and indexing date on electronic databases, search dates were from 1 year before the ending dates of the original searches through May 2010. For the report on cognitive function, the combination of generalist and specialty journals (based on the original review) resulted in a very small number of new titles; therefore, no restrictions in the search were needed because the goal of having a manageable number of titles to review was already achieved.

Studies were chosen based on the eligibility criteria of the original reports. Abstracts were screened by members of the respective EPC staff, and full-text articles were retrieved for all titles that appeared relevant; the articles were further screened for inclusion. Relevant systematic reviews also were included. For the RAND method, but not the modified Ottawa method, we added all articles cited by the domain experts.

2.4. Data extraction and data analysis

Data from the relevant articles were extracted by members of the respective EPCs. Each EPC conducted the comparison of the RAND method and the modified Ottawa method for the same evidence reviews. The same EPC staff participated in both applications; thus outcome assessors were not blinded. When implementing the two methods in parallel, we made every effort to ignore the findings of one method in evaluating the conclusions of the other.

All findings were compiled and analyzed together. The RAND assessments were compared with the modified Ottawa assessments for each key question and outcome, and a kappa statistic was calculated for each. We combined all the RAND signals, “definitely,” “probably,” and “possibly,” out of date into a single category, “out of date,” and compared the conclusions between methods within clinical topic area and across topics. The kappa statistic can be considered the chance-corrected proportional agreement. We used the Landis and Koch [14] interpretation of values of kappa to determine the level of agreement: <0, poor agreement; 0–0.20, slight agreement; 0.21–0.4, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.0, almost perfect agreement.

2.5. Role of the funding source

The AHRQ and the Office of Dietary Supplements participated in the formulation of the research methods but did not participate in conducting the analysis or in the preparation, conclusions, or approval of the manuscript for publication.

3. Results

The search results and article flow for the three evidence reports (neurological disorders/cognitive function [10], cancer [11], and risk factors for CVD [12]) are summarized in Fig. 1. First we present the topic-specific findings from implementing the RAND and modified Ottawa methods; then we provide a summary comparison between the two methods.

3.1. Effects of omega-3 FAs on cognitive function

The original report examined the effects of omega-3 FAs on treating or preventing dementia, maintaining cognitive function during aging, and preventing Parkinson’s disease and multiple sclerosis [10]. It included 12 studies that reported data on five outcomes/key questions. Because the original review did not include any meta-analyses, only the RAND and Ottawa qualitative methods were applied. From 89 newly identified full-text articles reviewed, 26 met eligibility criteria.

For the RAND method, invitations were sent to 25 experts, of whom five agreed to complete the forms. The invited experts included those who served as experts or reviewers of the original reviews, peer reviewers of the reports, and other subject matter experts known or recommended to us. Many of the experts, including some who did respond, indicated that they had little knowledge about the current state of research related to this field (this comment was shared by invited experts for all three topics). Only one of the five articles suggested by experts met the eligibility criteria, and it was also captured by the focused literature search. The other four articles either did not report outcomes of interest or were narrative review articles.

The RAND and modified Ottawa updating signals had an absolute agreement of 100% and kappa of 1.0 for the five key questions. Specifically, key questions 1–3 were determined to be out of date, and key questions 4 and 5 were considered up to date by both methods (Table 3A).

3.2. Effects of omega-3 FAs on cancer

The original report examined the effects of omega-3 FAs on incident cancer and recovery from surgery to treat gastrointestinal cancer [11]. It included 25 studies that reported data on 22 cancer outcomes/key questions. The original review included meta-analyses for three of the analyzed outcomes. From 67 newly identified full-text articles reviewed, 47 met eligibility criteria.

For the RAND method, invitations were sent to 13 experts, of whom seven agreed to complete the forms. The experts added 15 unique studies from journals outside the scope of the literature search. The RAND method produced a signal suggesting the probable or definite need for updating for four of 22 key questions.

The Ottawa quantitative approach (with no modification) was applied to three meta-analyses. No quantitative signal was found after adding all seven new trials. When we reviewed the remaining new studies for the modified Ottawa qualitative or other signals, we found one positive qualitative signal.

The RAND and modified Ottawa updating signals agreed for 20 of 22 key questions/outcomes (Table 3B), resulting in a kappa of 0.62 (95% confidence interval [CI] 0.12–1.0) for all key questions.

3.3. Effects of omega-3 FAs on cardiovascular risk factors

Following the structure of the original report [12], we treated the evaluation of fish oils and alpha-linolenic acid (ALA) as two separate systematic reviews for the current analysis. Together, the original report [12] and accompanying journal articles [15,16] included meta-analyses for nine of 26 analyzed CVD risk factors with fish oil and no metaanalyses for 24 CVD risk factors with ALA. The literature search yielded 54 new primary studies and six new systematic reviews.

For the RAND method, invitations were sent to 15 experts, of whom six completed the forms. The experts provided 12 additional articles from journals outside the scope of the literature search. These articles were not considered because they did not meet the eligibility criteria. The RAND method found signals for updating the report for 10 of 26 fish oil-related outcomes and 11 of 24 ALA-related outcomes. These conclusions were based mostly on new studies that were discordant with the original reports and/or the opinions of a minority of experts. Of note, though, the evidence provided by the experts did not always support their conclusions.

Several outcomes of one very large trial of fish oil (18,645 participants) were discordant with those of almost all other trials (both old and new) [17]. Thus, using the Ottawa quantitative method (with no modification), a quantitative signal was found after adding this single trial to the meta-analyses for six of seven outcomes. However, because of incomplete reporting, we had to estimate the standard errors of the net differences from this trial, and our estimates of the *P*-values of the net differences were discordant with the study conclusions for two outcomes (we estimated statistically significant CIs, but they reported no significant difference). This disparity affected the meta-analyses, resulting in statistically

significant, discordant meta-analysis estimates, despite qualitative agreement across studies of no effect. No modified Ottawa quantitative signal was found for two other eligible outcomes after the addition of 10 and 4 new trials, respectively.

With the qualitative approach of the modified Ottawa method, for all but two of 26 outcomes with fish oil, pivotal new trials either agreed with the original report or did not exist. For ALA, seven of 24 outcomes with no trial data in the original report had a small number of new small trials from the update search, resulting in classification as signals.

The RAND and modified Ottawa updating signals agreed for 18 of 26 outcomes (Table 3C) and for 22 of 24 outcomes (Table 3D), resulting in kappas of 0.30 (95% CI 0–0.70) and 0.83 (95% CI 0.60–1.0) for all the key questions related to the role of fish oil or ALA and cardiovascular risk factors, respectively.

3.4. Comparing results of the RAND and modified Ottawa methods across three reports

Across the three reports with four systematic reviews, the range of agreement varied from fair, with a kappa of 0.30, to almost perfect agreement, with a kappa of 1.0. Overall across all 77 conclusions, agreement was classified as “substantial” (kappa = 0.64, 95% CI 0.45–0.83) (Table 4).

4. Discussion

In this pilot study, we compared the RAND and modified Ottawa methods for identifying signals for the need to update systematic reviews and found “substantial” or better agreement in three of four systematic reviews. Overall agreement between methods was “substantial,” which supports the use of either method.

Several challenges have been identified in applying the two methods. First it was difficult to implement the RAND method when evaluating outcomes with sparse data. Specifically, we had difficulty determining whether outcomes with zero to two studies in the original report were outdated when there were one or two new studies with small or nonsignificant results. We found that we came to different conclusions each time we reviewed the new evidence. Second, further consideration should be given regarding what to conclude when a single expert votes that a topic is out of date and how to interpret experts’ votes that a topic is outdated if no supporting evidence (such as the citations to support the outdated claims) is provided. Analogously, further consideration needs to be given to the situation where a single trial provides a quantitative signal, where experts did not judge this new evidence to be a signal for updating, and the researchers who conducted the original systematic review also judged the new evidence to be inconclusive in changing the original finding. Third, although it would have been desirable to compare the resources involved in applying each method, we found it impossible to do that because each EPC was conducting the comparison of the RAND method and the modified Ottawa method for the same evidence reviews and the same EPC staff was participating in both applications. The RAND method imposed an additional workload on outside domain experts, but we did not attempt to quantify this. Qualitatively, we note that the Ottawa approach alone would involve less work than the RAND method if one of the larger RCTs triggered a quantitative signal, as occurred with

multiple outcomes for fish oil in the CVD report. Conversely, when no quantitative signal is found, even after adding all seven new trials for the existing three meta-analyses, as seen in the assessment for the cancer review, the Ottawa method might take more work than the RAND method.

This test case has several limitations. First, we were able to evaluate only three reports with four systematic reviews, all about the same intervention. Even with a small sample, we found some heterogeneity, with one analysis having only a fair agreement ($\kappa = 0.3$) and three having substantial agreement. The “fair” agreement between the two methods (for the fish oil report) can be explained by the confluence of an atypically large new trial driving the Ottawa quantitative signals and the experts reporting that some report signals were out of date without supporting evidence, as discussed earlier. Future evaluations might identify whether or not there are factors, which predict which topics will have poor agreement. Second, the same EPC staff implemented the two update signal methods in parallel without blinding; thus it is possible that the signals from one method could have influenced the judgment of update signals using the other method. Moreover, the judgments of update signals are largely subjective. It should be noted that the Ottawa quantitative signals are less susceptible to subjectivity because the judgments of quantitative signals were explicitly defined. Finally, we cannot compare the predictive validity of the RAND and modified Ottawa methods as no actual updates of the original reviews have been conducted. Such a predictive validity analysis will need to wait until reports assessed for signals are actually updated. In the future, after several updates have been performed on evidence reviews that have been analyzed for signals on wide ranges of clinical topics, it will be useful to analyze whether the modified Ottawa and RAND methods, or a combination of the two, accurately predict whether the conclusions of the updated systematic review differ from those of its predecessor.

In the process of implementing the two update signal methods in parallel, we ignored the findings of one method in evaluating the conclusions of the other. In particular, in conducting the modified Ottawa method, we did not include the new studies supplied to us by the domain experts. This artificial approach highlighted that there is not an either/or decision to be made as to which update signal method should be applied in the future. It seems logical that a hybrid approach may be most reasonable, using input from domain experts and searching for pivotal trials or meta-analytic evidence of quantitative signals. Using a hybrid approach may also mitigate both overly optimistic expert interpretation of the new data as a need to update (using the RAND method) and overinterpretation of large, highly precise studies without consideration of potential methodological flaws, applicability issues, or the clinical significance of effects (using the Ottawa method). Additional consideration should be given to the utility of other kinds of signals, such as the continuing use and importance of the original systematic review, the continuing use of the interventions assessed in the review, and whether there is an opportunity for the updating of the review to lead to a change in practice.

Factors that may influence the choice of method, although not tested explicitly by us, may include the following:

- Level of expert engagement in the research topic (low levels favor the modified Ottawa method),
- Quality and variation in study designs found in the new evidence (low-quality or variable study designs favor the RAND method, as the modified Ottawa method is designed with high-quality trials in mind),
- Desire for considering absolute levels of prior evidence, rather than only relative levels (high desire favors the RAND method, which allows more subjective application of updating signals; the modified Ottawa method's relative change signals do not take into account whether the original review included either very few or many studies),
- Desire for a transparent, consistent signaling method that maximizes interrater reliability (high desire probably favors the modified Ottawa method because signals have less flexibility).

In conclusion, our results support the use of either method; in general the RAND and modified Ottawa methods provide similar signals for the possible need to update systematic reviews. A decision to update a systematic review also might be informed by the application of both methods, with the results compared to provide additional validation or highlight areas of disagreement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The principal investigators and research team would like to acknowledge the guidance provided by the Task Order Officer, Margaret Coopey; the Office of Dietary Supplements' Director, Paul Coates; the assistance of the Ottawa EPC librarian, Becky Skidmore; administrative assistants, Patty Smith; and the careful work of the reviewers.

Role of funding source: AHRQ had input into the general concept of the study but no role in the data collection and analysis or interpretation of the results.

References

1. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *J Clin Epidemiol.* 2010; 63:491–501. [PubMed: 19540721]
2. Garrity C, Tsertsvadze A, Tricco AC, Sampson M, Moher D. Updating systematic reviews: an international survey. *PLoS ONE.* 2010; 5:e9914. [PubMed: 20376338]
3. The Cochrane Collaboration. Maintaining your review. 2002 Available at <http://www.cochrane-net.org/openlearning/html/mod19-2.htm>.
4. French S, McDonald S, McKenzie J, Green S. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC Med Res Methodol.* 2005; 5:33. [PubMed: 16225692]
5. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? *BMJ.* 2001; 323:155–157. [PubMed: 11463690]
6. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, et al. Validity of the Agency for Healthcare Research and Quality Clinical Practice Guidelines: how quickly do guidelines become outdated? *JAMA.* 2001; 286:1461–1467. [PubMed: 11572738]

7. Shekelle, PG.; Newberry, SJ.; Maglione, M.; Shanman, R.; Johnsen, B.; Carter, J., et al. Assessment of the need to update comparative effectiveness reviews: report of an initial rapid program assessment (2005–2009) (prepared by the Southern California Evidence-based Practice Center). Rockville, MD: Agency for Healthcare Research and Quality; 2009. Available at: www.effectivehealthcare.ahrq.gov/ehc/products/125/331/2009_0923UpdatingReports.pdf.
8. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, et al. When and how to update systematic reviews. *Cochrane Database Syst Rev*. 2008:MR000023. [PubMed: 18254126]
9. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007; 147:224–233.
10. Maclean, C.; Issa, A.; Newberry, S.; Mojica, W.; Morton, S.; Garland, R. AHRQ Publication Number 05-E011-1. Rockville, MD: Agency for Healthcare Research and Quality; 2005. Effects of omega-3 fatty acids on cognitive function with aging, dementia, and neurological diseases. Summary, evidence report/technology assessment: number 114. [11-30-2010]
11. Maclean, C.; Newberry, S.; Mojica, W.; Issa, A.; Khanna, P.; Lim, Y. AHRQ Publication Number 05-E010-1. Rockville, MD: Agency for Healthcare Research and Quality; 2005. Effects of omega-3 fatty acids on cancer. Summary, evidence report/technology assessment: number 113. [11-30-2010]
12. Balk, E.; Chung, M.; Lichtenstein, A.; Chew, P.; Kupelnick, B.; Lawrence, A., et al. AHRQ Publication No. 04-E010-2. Rockville, MD: Agency for Healthcare Research and Quality; 2004. Effects of omega-3 fatty acids on cardiovascular risk factors and intermediate markers of cardiovascular disease. Evidence report/technology assessment no. 93.
13. Balk EM, Horsley TA, Newberry SJ, Lichtenstein AH, Yetley EA, Schachter HM, et al. A collaborative effort to apply the evidence-based review process to the field of nutrition: challenges, benefits, and lessons learned. *Am J Clin Nutr*. 2007; 85:1448–1456. [PubMed: 17556679]
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
15. Balk EM, Lichtenstein AH, Chung M, Kupelnick B, Chew P, Lau J. Effects of omega-3 fatty acids on serum markers of cardiovascular disease risk: a systematic review. *Atherosclerosis*. 2006; 189:19–30. [PubMed: 16530201]
16. Balk EM, Lichtenstein AH, Chung M, Kupelnick B, Chew P, Lau J. Effects of omega-3 fatty acids on coronary restenosis, intima-media thickness, and exercise tolerance: a systematic review. *Atherosclerosis*. 2006; 184:237–246. [PubMed: 16084516]
17. Oikawa S, Yokoyama M, Origasa H, Matsuzaki M, Matsuzawa Y, Saito Y, et al. Suppressive effect of EPA on the incidence of coronary events in hypercholesterolemia with impaired glucose metabolism: sub-analysis of the Japan EPA Lipid Intervention Study (JELIS). *Atherosclerosis*. 2009; 206:535–539. [PubMed: 19447387]

What is new?

Key finding

- Two methods to identify signals for the need to update systematic reviews substantively agreed in a study analyzing three evidence reports on omega-3 fatty acids.

What this adds to what was known?

- Our test case supports the use of either method; in general they provide similar signals for the possible need to update systematic reviews.

What is the implication, what should change now?

- Agencies charged with developing and maintaining systematic reviews can use either method for surveillance for signals of the need to update.
- Several factors may influence the choice of method, such as level of expert engagement in the research topic, quality and variation in study designs found in the new evidence, desire for considering absolute levels of prior evidence, rather than only relative levels, and desire for a transparent, consistent signaling method that maximizes interrater reliability.

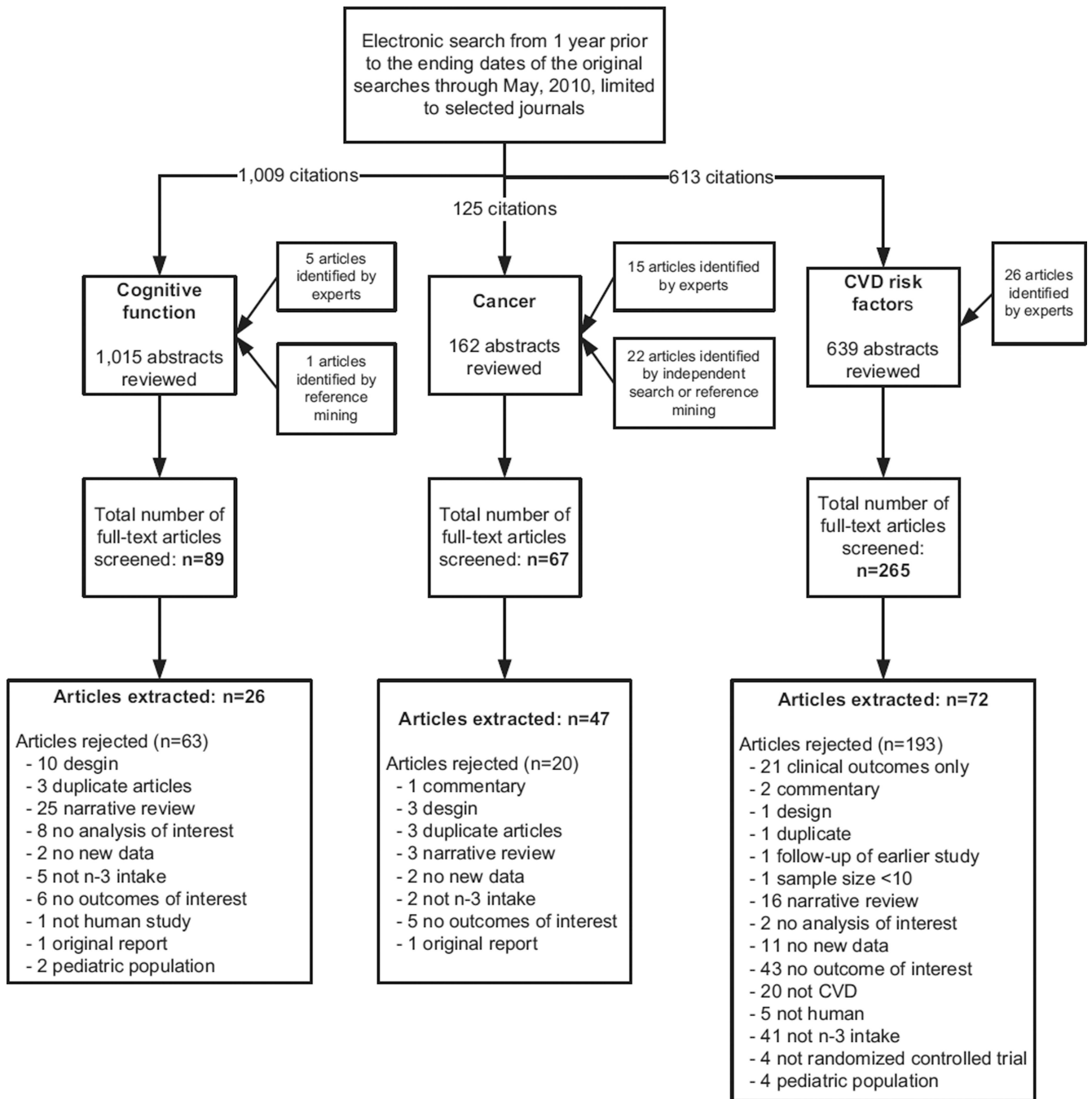


Fig. 1. Search results and article flow for each of the three omega-3 fatty acid evidence reports.

Table 1

The RAND method for identifying signals for the need for an update

Four-category scheme	Operational definitions
Definitely out of date	Original conclusion is out of date: if we found new evidence that rendered the conclusion out of date or no longer applicable, we classified the conclusion as out of date. Recognizing that our literature searches were limited, we reserved this category only for situations where a limited search would produce prima facie evidence that a conclusion was out of date, such as the withdrawal of a supplement from the market.
Probably out of date	Original conclusion is probably out of date, and this portion of the original report may need updating: if we found substantial new evidence that might change the conclusion, and/or most responding experts judged the conclusion as having new evidence that might change the conclusion, then we classified the conclusion as probably out of date.
Possibly out of date	Original conclusion is possibly out of date, and this portion of the original report may need updating: if we found some new evidence that might change the conclusion, and/or a minority of responding experts judged the conclusion as having new evidence that might change the conclusion, then we classified the conclusion as possibly out of date.
Still valid	Original conclusion is still valid, and this portion of the original report does not need updating: If we found no new evidence or only confirmatory evidence and all responding experts judged the conclusion as still valid, we classified the conclusion as still valid.

Table 2

The modified Ottawa method for identifying signals for the need for an update

Type of signals	Signal code	Operational definitions
Qualitative signals of potentially invalidating changes in evidence	A1	Opposing findings: a pivotal trial ^a or meta-analysis (or guidelines), including at least one new trial that characterized the treatment in terms opposite to those used earlier.
	A2	Substantial harm: a pivotal trial ^a or meta-analysis (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making.
	A3	A superior new treatment: a pivotal trial ^a or meta-analysis (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.
Qualitative signals of major changes in evidence	A4 ^b	Important changes in effectiveness short of “opposing findings.”
	A5	Clinically important expansion of treatment.
	A6	Clinically important caveat.
	A7	Opposing findings from discordant meta-analysis or nonpivotal trial.
Quantitative signals	B1	A change in statistical significance (from nonsignificant to significant).
	B2	A change in relative effect size of at least 50 percent.
“Other” signals	n/a	“Other” signals were sought for key questions for which there were no prior meta-analyses or RCTs, for example, questions for which only large cohort or case—control studies were identified. The criteria included a major increase in the number of new studies or a new study with at least three times the number of participants as in previous studies. These criteria had to be adapted to account for situations, such as a large number of new, but smaller studies, when the studies in the original report had been large prospective cohort studies and the new studies were largely smaller nested case—control studies.

Abbreviations: RCT, randomized controlled trial; n/a, not applicable.

^aPivotal trial was defined as a trial published in a pivotal journal (see Methods) or trials published in nonpivotal journals with at least three times the number of participants as the previous largest trial.

^bThe original Ottawa method is silent on how to assign a signal when the original report had no trials addressing a specific question and there are new small trials. See Methods for how we modified this signal for the present study.

Table 3

Comparison of signals for updating reviews

Evidence reports	Ottawa positive	Ottawa negative	Total
A. Cognitive function review			
RAND definitely out of date	0	0	0
RAND probably out of date	3	0	3
RAND possibly out of date	0	0	0
RAND still valid	0	2	2
<i>Total</i>	3	2	5
Kappa = 1.00			
B. Cancer review			
RAND definitely out of date	1	0	1
RAND probably out of date	1	2	3
RAND possibly out of date	0	0	0
RAND Still valid	0	18	18
<i>Total</i>	2	20	22
Kappa = 0.62 (95% CI 0.12–1.0)			
C. Fish oil and cardiovascular risk factors review			
RAND definitely out of date	0	0	0
RAND probably out of date	0	1	1
RAND possibly out of date	4	5	9
RAND still valid	2 ^a	14	16
<i>Total</i>	6	20	26
Kappa = 0.30 (95% CI 0–0.70)			
D. ALA and cardiovascular risk factors review			
RAND definitely out of date	0	0	0
RAND probably out of date	7	0	7
RAND possibly out of date	2	2	4
RAND still valid	0	13	13
<i>Total</i>	9	15	24
Kappa = 0.83 (95% CI 0.60–1.0)			

Abbreviations: CI, confidence interval; ALA, alpha-linolenic acid.

Signals in agreement with each other are in bold.

^aThe results of one very large trial were discordant with those of almost all other trials. Our consensus was that this one outlier trial did not invalidate the original overall conclusions.

Table 4

Overall comparison and kappa statistic for all three reviews

Overall comparison	Ottawa positive	Ottawa negative	Total
RAND out of date	18	10	28
RAND still valid	2	47	49
Total	12	65	77

Signals in agreement with each other are in bold.

Kappa = 0.64 (95% CI 0.45–0.83).