Technical Review
Number 16

Updating Systematic Reviews

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0021

Prepared by:
University of Ottawa Evidence-based Practice Center, Ottawa, Canada

Investigators
Kaveh G. Shojania, M.D.
Margaret Sampson, M.L.I.S.
Mohammed T. Ansari, M.B.B.S., M.Med.Sc., M.Phil.
Jun Ji, MD, M.H.A.
Chantelle Garrity, B.A., D.C.S., M.Sc.(c)
Steve Doucette, M.Sc.
Tamara Rader, M.L.I.S.
David Moher, Ph.D.

AHRQ Publication No. 07-0087
September 2007
This report is based on research conducted by the University of Ottawa Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-02-0021). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.
Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this evidence report. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.gov.

Carolyn M. Clancy, M.D.  
Director  
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.  
Director, Center for Outcomes and Evidence  
Agency for Healthcare Research and Quality

Beth A. Collins Sharp, Ph.D., R.N.  
Director, EPC Program  
Agency for Healthcare Research and Quality

David Atkins, M.D., M.P.H.  
Chief Medical Officer  
Agency for Healthcare Research and Quality
Acknowledgments

We would like to acknowledge the guidance and expertise contributed to this project by the Technical Expert Panel members:

- David Atkins – Chief Medical Officer, Agency for Healthcare Research and Quality
- Paul Shekelle – RAND (USA)
- Evelyn P. Whitlock – Kaiser Permanente Center for Health Research (USA)
- Cynthia Mulrow – University of Texas, San Antonio and Annals of Internal Medicine (USA)
- Doug Altman – Center for Statistics in Medicine (UK)
- Martin Eccles – Center for Health Services Research, Newcastle University (UK)
- P.J. Devereaux – McMaster University Health Sciences Centre (CAN)

We would like to acknowledge the contributions of members of the University of Ottawa Evidence-based Practice Center and others. Many of these researchers made significant contributions to portions of the project.

- Lorri Puil, Director of Systematic Reviews, Chalmers Research Group, for support and supervision of research staff
- Keith O’Rourke and Nicholas J Barrowman, for statistical advice
- Andrea Tricco for guidance with survey methods
- Alexander Tsertsavadze, for early methodological discussions and assistance with screening
- Tanya Armour, for mentorship of team members
- Jessie McGowan, for guidance with search method development
- Raymond Daniel, for search assistance and document acquisition
- Alla E. Iansavichene, for search assistance and data quality management
- Mary Ocampo, for technical and administrative assistance
- Alison Jenkins, for development of the meta-analytic Excel worksheet
Structured Abstract

**Background:** Systematic reviews are often advocated as the best source of evidence to guide both clinical decisions and healthcare policy, yet we know very little about the extent to which they require updating.

**Objectives:**
- To estimate the average time to changes in evidence sufficiently important to warrant updating systematic reviews (referred to as the survival time) and to identify any characteristics that increase or decrease these survival times.
- To determine the performance characteristics of various surveillance protocols to identify important new evidence.
- To assess the utility of rates and patterns of growth for evidence within clinical areas as predictors of updating needs.
- To establish typical timeframes for the production and publication of systematic reviews in order to assess the extent to which they impact survival time (e.g., whether or not delays in the peer review and publication processes substantially shorten the time in the public domain before new evidence requires updating of a given systematic review).
- To characterize current updating practices and policies of agencies that sponsor systematic reviews.

**Design:** Survival analysis for a cohort of 100 quantitative systematic reviews that were indexed in *ACP Journal Club* with an accompanying commentary; supplementary sample of Cochrane reviews meeting the same criteria and AHRQ evidence reports; internet-based survey of agencies that sponsor or undertake systematic reviews.

**Sample:** Eligible reviews evaluated the clinical benefit or harm of a specific (class of) drug, device, or procedure, were originally published between 1995 and 2005, and included at least one quantitative synthesis result in the form of an odds ratio, relative risk, risk difference, or mean difference. For the survey of updating policies and practices, we contacted 22 organizations that are well-known to produce or fund systematic reviews (including 12 AHRQ Evidence-based Practice Centers).

**Data sources:** Systematic reviews indexed in ACP J Club and eligible new trials identified through five search protocols.

**Measurements:** Quantitative signals for updating consisted of changes in statistical significance or a relative change in effect magnitude of at least 50 percent involving one of the primary outcomes of the original systematic review or any mortality outcome. These signals were assessed by comparing the original meta-analytic results with updated results that included eligible new trials. Qualitative signals included substantial differences in characterizations of effectiveness, new information about harm, emergence of superior alternative treatments, and important caveats about the previously reported findings that would affect clinical decisionmaking.
The primary outcome of interest was the occurrence of either a qualitative or quantitative signal for updating the original systematic review. We also assessed the occurrence of a signal for updating within 2 years of publication, as some sources (e.g., The Cochrane Library) currently recommend updating systematic reviews every two years.

The survey measured existing updating policies, current strategies in use, and additional perceptions related to the updating process from the 18 organizations that responded.

**Results:** The cohort of 100 systematic reviews included a median of 13 studies (inter-quartile range: 8 to 21) and 2663 participants (inter-quartile range: 1281 to 8371) per review. A qualitative or quantitative signal for updating occurred for 57 systematic reviews. Median survival free of a signal for updating was 5.5 years (95% confidence interval [CI]: 4.6-7.6), but in 23 cases (95% CI: 15% to 33%), a signal for updating occurred in less than 2 years, and in 15 cases (95% CI: 9% to 24%) the signal occurred in less than 1 year. In 7 cases (95% CI: 3% to 14%), a signal had already occurred at the time of publication of the original review. Shorter survival was associated with cardiovascular medicine (hazard ratio of 3.26, 95% CI: 1.71 to 6.21; p =0.0003), heterogeneity in the original review (hazard ratio of 2.23, 95% CI: 1.22 to 4.09; p =0.01), and having a new trial larger than the previous largest trial (hazard ratio of 1.08, 95% CI: 1.02 to 1.15; p =0.01). Systematic reviews with more than the median of 13 included studies had increased survival (hazard ratio of 0.55; 95% CI: 0.31 to 0.98; p =0.04). No feature of the original review significantly predicted a signal for updating occurring within 2 years of publication.

Median time from the final search date to indexing 1.4 years (inter-quartile range: 0.96-2.0 years). Lags from search to publication were shortest for Cochrane reviews (median 0.6 years, inter-quartile range: 0.42-1.25) and longest for journal reviews (median 1.3 years; inter-quartile range: 0.84-1.77), with technical reports falling in between (median 1.1 years; inter-quartile range: 0.87-1.42) (Kruskal Wallis \( \chi^2 \) 11.24, p =0.004).

Of the five search protocols tested for their effectiveness in identifying eligible new trials, the combination with the highest recall and lowest screening burden were the strategy that used the PubMed Related Articles feature (applied to the three newest and three largest trials included in the original review) and the strategy involved submitting a subject search (based on population and intervention) to the Clinical Query filter for therapy. This combination identified most new signaling evidence with median screening burden of 71 new records per review.

For the survey of organizations involved in producing or funding systematic reviews, we received responses from 19 (86%) of the 22 organizations contacted. Approximately two thirds (68%) of respondents identified themselves as producers of systematic reviews and an additional 21% identified themselves as both funders and producers of systematic reviews. Only two respondents (11%) characterized themselves solely as funders of systematic reviews.

Approximately 80% of respondents characterized the importance of updating as ‘high’ or ‘very high’, although 68% acknowledged not having any formal policies for updating in place. Approximately two thirds (13/19; 68%) of respondents reported that over 20% of the reviews they commission or produce are out of date, and 32% respondents (6/19) reported that at least 50% of their reviews were out of date. Barriers to updating identified by respondents included lack of appropriate methodologies, resource constraints, lack of academic credit, and limited publishing formats. The majority of the sample (16/19; 84%) indicated they ‘somewhat’ to
‘strongly’ favor the development of a central registry, analogous to efforts within the clinical trials community, to coordinate updating activities across agencies and review groups.

**Conclusions:** In a cohort of high quality systematic reviews directly relevant to clinical practice, signals for updating occurred frequently and within relatively short timelines. A number of features significantly affected survival, but none significantly predicted the need for updating within 2 years.

Currently, definitive methods about the frequency of updating cannot be made. Blanket recommendation such as every two years will miss a substantial number of important signals for updating that occur within shorter time lines, but more frequent updates will expend substantial resources. Methods for identifying reviews in need of updating based on surveillance for new evidence hold more promise than relying on features of the original review to identify reviews likely to need updating within a short time, but such approaches will require further investigation. Several of the methods tested were feasible, yielding good recall of relevant new evidence with modest screening burdens.

The majority of organizations engaged in the funding or production of systematic reviews view the importance of updating systematic reviews as high to very high. Despite this recognition, most organizations report having no formal policy in place for updating previous systematic reviews. Slightly less than half of organizations performed periodic literature searches to identify new evidence, but searching frequencies varied widely, from monthly to every two years.

If systematic reviews are to achieve their stated goal of providing the best evidence to inform clinical decision making and healthcare policy, issues related to identifying reviews in need of updating will require much greater attention. In the meantime, publishers of systematic reviews should consider a policy of requiring authors to update searches performed over 12 months prior to submission. And, users of systematic reviews need to recognize that important new evidence can appear within short timelines. When considering the results of a particular systematic review, users should search for more recent reviews or trials to see if any exist and determine if the results are consistent with the previous review.
# Contents

Technical Review .............................................................................................................................. 3

Chapter 1. Introduction ....................................................................................................................... 3
  Overview ......................................................................................................................................... 3
  Background ................................................................................................................................. 3
  Key Questions Addressed in This Report .................................................................................. 4

Chapter 2. Methods ............................................................................................................................. 5
  Study Identification ......................................................................................................................... 5
  Search Strategy ............................................................................................................................ 5
  Eligibility Criteria .......................................................................................................................... 5
  Cohort Selection Process .......................................................................................................... 6
  Data Collection ............................................................................................................................. 7
    I. Signals for Updating and Survival Analysis for the Cohort of 100 Systematic Reviews .......... 7
    II. Detection of Quantitative Signals for Updating .................................................................. 11
    III. Publication Time Lags ...................................................................................................... 13
    IV. Growth of the Literature by Clinical Area ....................................................................... 13
    V. Survey of Organizations Engaged in Funding or Production of Systematic Reviews .......... 14
  Analysis ......................................................................................................................................... 14
    Survey analysis ......................................................................................................................... 15

Chapter 3. Results ............................................................................................................................... 17
  Results of Literature Search and Cohort Screening ....................................................................... 17
    Assessment of the new evidence ............................................................................................... 17
  Characteristics of Included Studies ............................................................................................. 17
    Composition of the cohort ....................................................................................................... 17
    Signals for updating ............................................................................................................... 20
    Survival analysis .................................................................................................................. 20
    Directions of changes in evidence and expected impact on practice .................................... 26
  Search Performance .................................................................................................................... 27
    Adequacy of MEDLINE® coverage for surveillance ............................................................... 29
  Time Lags in the Production and Publication of Systematic Reviews ...................................... 30
  Publication Velocity
  Policies and Practices of Agencies or Organizations that Fund or Conduct Systematic Reviews
    Systematic Reviews ................................................................................................................. 33
    Respondents ............................................................................................................................. 33
    Main findings ........................................................................................................................... 33

Chapter 4. Discussion ......................................................................................................................... 39
  Practical Implications .................................................................................................................. 38
    For users of systematic reviews ......................................................................................... 39
Tables

Table 1. Characteristics of the Cohort of 100 Systematic Reviews ............................................... 19
Table 2. Frequency of the Different Types of Signals for Updating ............................................... 22
Table 3. Univariate Survival Analysis .......................................................................................... 24
Table 4. Multivariate Analysis of Hazards .................................................................................... 25
Table 5. Changes in Certainty and Expected Impacts on Practice Associated with Signals for Updating. .......................................................... 29
Table 6. Recall of Signaling Evidence by the Surveillance Searches ............................................. 31
Table 7. Linear Fit And Rate Of Growth By Clinical Area, Doubling Time In Years From Various Starting Years Assuming Linear Growth. ......................................................... 34
Table 8. Monitoring Strategies .................................................................................................... 35
Table 9. Factors that Impact on Determining “When” to Update .................................................. 36
Table 10. Methods/Procedures .................................................................................................. 38
Table 11. Major/Moderate Benefits to Harmonization .................................................................. 38

Appendixes

Appendix A. Definitions and Criteria for Signals for Updating
Appendix B: Examples of Systematic Reviews with Qualitative or Quantitative Signals for Updating
Appendix C: Sample Subject Searches
Appendix D: Sample Assessment Worksheet
Appendix E: Searches by Clinical Area
Appendix F: Growth of the Literature By Clinical Area
Appendix G: Survey Instrument

Technical Review
Chapter 1. Introduction

Overview

Systematic reviews are being published with increasing frequency. One recent estimate is that 2500 new systematic reviews are published annually.¹ This high volume reflects the several key roles played by systematic reviews: synthesizing for clinicians the evidence addressing a given topic; providing the foundation for the development of clinical practice guidelines; and informing cost-effectiveness analyses and policy decisions. In addition, some granting agencies now require that researchers include systematic reviews in grant applications to support the rationale for proposed new research. Fulfilling these important roles requires that systematic reviews be up to date. However, almost no empiric data indicate the extent to which systematic reviews require updating or the intervals at which updates should be performed. Given the paucity of data on this important topic, and with funding through the Evidence-based Practice Center, we set out to generate data to help inform optimal approaches to updating.

Background

The annual publication of systematic reviews has increased dramatically in recent years,²,² with an estimated 2500 new systematic reviews published per year.¹ The Cochrane Collaboration sets a goal of updating its reviews every 2 years, with the result that 38% of new Cochrane reviews represent updates of previous reviews (typically conducted by the same authors).¹ By contrast, only 2% of systematic reviews published in all other journals represent updates of previous reviews (whether conducted by the same authors or not).¹ Currently we have no way of knowing if either of these numbers adequately matches the true need for maintaining the currency of published systematic reviews.

Updating a previous systematic review requires resources. Even when the same authors update the review, carrying out the search and screening processes for potentially eligible new articles, abstracting new articles that are identified, and analyzing the results all take time. Agencies that commission and host systematic reviews thus face the challenge of how best to allocate resources between funding new reviews and supporting the maintenance of existing reviews. Even the process of determining if an aging review still represents a valid, even if no longer completely up to date, synthesis that can safely be left in the public domain, or new evidence alters the findings of the review to such an extent that withdrawing or archiving the review represents the preferred choice.

The goals of this effort are to determine the extent to which systematic reviews require updating and to ascertain approaches towards updating by organizations engaged in the funding or production of systematic reviews. This work builds upon and complements several pieces of work in which members of this team have participated over the past 24 months, including development of the first formal definition of updating to appear in the literature,³ and a systematic review of existing methodologies for updating.⁴ The evidence base proved to be quite limited, with a small body of literature on cumulative meta-analysis and the volume of new evidence needed to overturn previous meta-analytic results,⁵,⁶ one evaluation of the shelf-life of
17 clinical practice guidelines produced for the Agency for Healthcare Research and Quality, and some suggested approaches to updating from the Cochrane Collaboration.

Given this paucity of literature informing approaches to updating, we set out to evaluate how soon systematic reviews require updating, and how best to detect the need for updating through empirical study of a cohort of quantitative systematic reviews. We also sought to determine current practices with respect to updating through consultation with agencies involved with funding or producing systematic reviews.

**Key Questions Addressed in This Report**

This evidence report aims to provide empiric data that address the following questions about updating systematic reviews.

1. How quickly do systematic reviews become out of date (i.e., what is the average 'shelf life' or 'survival time' of systematic reviews)?
2. Do any features of a given systematic review (including characteristics related to the content area, features of the included studies, and the nature of the results) increase or decrease ‘survival’ time?
3. What impact do publication time lags have on survival times for systematic reviews and what strategies can maximize the currency of systematic reviews at the time of publication?
4. Is the pattern of growth of evidence (e.g., the velocity of trial publication) within clinical areas predictive of the need to update?
5. Of agencies or organizations that fund or conduct systematic reviews, what are their existing policies or practices regarding updating systematic reviews?

Questions 1 to 4 are the main focus of this evidence report. They are addressed through identification and evaluation of new evidence pertaining to each of 100 systematic reviews, with an augmented sample of Cochrane reviews and AHRQ evidence reports for Question 3. Methods and main finding are described here in this extended overview. Each question is fully reported in a journal manuscript. The fifth question was addressed through a survey of 9 agencies as well as the EPCs, who also responded to several additional questions of particular interest to the AHRQ’s EPC Program. The survey served as a pilot for a larger survey not conducted under the auspices of AHRQ. The material from the pilot will be reported in a separate manuscript. Methods and major results of the pilot survey are presented here; the larger survey will be reported in a subsequent manuscript.
Chapter 2. Methods

Study Identification

Search Strategy

The first four questions are explored through a cohort of 100 systematic reviews identified through a search of ACP Journal Club database on Ovid, undertaken January 31, 2006. The search to identify candidates screened for inclusion in the cohort was:

1. review$.ti.
2. meta-analy$.mp.
3. data sources.ab.
4. (search$ or MEDLINE®).ab.
5. or/1-4
6. limit 5 to articles with commentary

Additional Cochrane reviews included for Question 3 were identified through the same search. Additional AHRQ reports used for Question 3 were identified through PubMed® with the query "Evid Rep Technol Assess (Summ)"[Journal:__jrid21544]. Searches were undertaken April 10, 2006.

Eligibility Criteria

The time to important changes in evidence might vary depending on a number of factors, including the type of question posed by the original review (e.g., therapeutic, diagnostic, prognostic, or health policy), the type of studies included (e.g., randomized controlled trials, observational studies), and whether or not the systematic review provided quantitative synthesis. In the interest of reducing potential sources of variation, we focused on systematic reviews that evaluated the clinical benefit or harm of a specific (class of) drug, device, or procedure and provided quantitative synthesis that included a point estimate and 95% confidence interval for at least one clinical outcome (disease endpoint, functional status, mortality) or established intermediate outcome (e.g., blood pressure, glycemic control, standard instrument for measuring disease activity, such as a depression scale). We excluded evaluations of alternative and complementary medicines, as well as educational and behavioral interventions.

Further eligibility requirements were as follows:

- Publication from 1995 to 2005 (but with search date no later than Dec 31, 2004 to ensure at least one full year for new evidence to appear)
- Reporting of at least one conventional meta-analytic estimate of treatment benefit or harm. We excluded individual patient data meta-analyses, meta-regressions, and indirect meta-analyses because of the difficulty of determining whether or not data from new trials would alter previous quantitative results.
• Included at least one randomized controlled trial; other eligible designs were restricted to quasi-randomized or controlled clinical trials (CCTs).
• Meta-analytic outcomes reported in the form of a relative risk, odds ratio, or absolute risk difference for binary outcomes and weighted mean differences for continuous outcomes. We excluded standardized effect sizes to avoid the complexity of assessing candidate new data reported using different outcome scales to determine if they would have met the authors’ criteria for incorporation into the standardized effect measure in the original review.

We used as our sampling frame systematic reviews that were selected for commentaries in *ACP Journal Club,* a bimonthly publication of the American College of Physicians that aims “to select from the biomedical literature articles that report original studies and systematic reviews that warrant immediate attention by physicians attempting to keep pace with important advances in internal medicine.” 11 The article selection process involves “reliable application of explicit criteria for scientific merit, followed by assessment of relevance to medical practice by clinical specialists.” Moreover, systematic reviews indexed in *ACP Journal Club* must meet specific quality criteria. Thus, choosing this sampling frame allowed us to identify systematic reviews of reasonable quality (or better) that are directly relevant to clinical practice.

**Cohort Selection Process**

Each record identified through the search of *ACP Journal Club* was screened for eligibility on the basis of title and abstract by 2 reviewers. Records with consensus in favor of eligibility were promoted, where final confirmation of eligibility was made based on the full report. Records were screened in alphabetical order by first author until 100 eligible reviews were identified. We chose a sample size of 100 to balance the practical issue of time required to ascertain the need for updating for each review in the cohort with power considerations, such as the expected width of confidence intervals given a denominator of 100 and the ability to evaluate predictive models of the need for updating with at least 3 to 5 potential predictors in the models. Of the 100 total reviews, we set the maximum number of Cochrane reviews was to 30. We chose to limit the number of Cochrane reviews, as evidence suggests that they differ in important respects from other systematic reviews in the peer review literature on the basis of style and possibly on topic coverage.1

A supplemental sample was formed for question 3 as additional eligible reviews beyond the 100 had been identified, and because data extraction was quick and a larger cohort would facilitate comparisons between report types, these additional reports were included in the cohort for question 3. Few eligible HTA reports were identified through *ACP Journal Club* so Evidence Reports that were otherwise eligible were added to permit comparisons of production milestones between HTA reports undertaken by AHRQ and other types of reviews.

When an eligible review was an explicit update of an earlier review (e.g., in the case of Cochrane reviews, which are updated and reissued periodically as a matter of policy), we used the earliest version in the time frame of 1995-2005. Similarly, when more than one review on the same topic was identified, only the earliest was included, to avoid double counting the same changes in evidence (or lack thereof).
We abstracted data on primary outcomes for each systematic review. To qualify as primary outcomes, we required that authors use the words “primary” or “main” and that they identify no more than 3 such outcomes (i.e., we regarded identification of more than 3 “primary” outcomes as inconsistent with the concept of primary outcome). For reviews that did not identify primary outcomes, we selected outcomes in the order in which their results were presented, including up to 4 efficacy outcomes and up to 2 harm outcomes. Eligible outcomes were clinical outcomes (disease endpoint, functional status, mortality) or established intermediate outcome (e.g., blood pressure, glycemic control, standard instrument for measuring disease activity, such as a depression scale). Each must have provided an eligible quantitative synthesis in the formats noted above (relative risk, odds ratio, or absolute risk difference for binary outcomes and weighted mean differences for continuous outcomes).

Data Collection

The several questions reported here (detecting updating signals for the cohort of 100 quantitative systematic reviews, publication time lags for a larger cohort of 148 reviews, analysis of the patterns of growth in evidence in different clinical areas, and the survey of organizations involved in systematic review work regarding updating practices) involved different data collection methods. These details are presented in sections for each project.

I. Signals for Updating and Survival Analysis for the Cohort of 100 Systematic Reviews

Data extraction from the cohort reviews. For each of the 100 systematic reviews, we characterized the type of intervention (drug, device, or procedure), the numbers of included trials and participants, methodological features, such as the presence of heterogeneity or publication bias, descriptions of reported outcomes and identification of those explicitly identified as ‘primary’ or ‘main,’ the meta-analytic results for each outcome, and excerpted quotations of the authors’ characterizations of these results and their interpretation of them.

We also classified all reviews into a clinical area. For reviews published in print journals, we primarily based this classification on the ISI classification of the clinical area of the journal in which the review appeared. For reviews published in general journals, Cochrane reviews, and HTA reports, we considered the specialty journals for which the review would have been most suitable. In the case of Cochrane reviews, we also based the classification of clinical content area on the review group that carried out the work (e.g., the Cochrane Musculoskeletal Group, the Cochrane Metabolic and Endocrine Disorders Group). For other types of reviews (e.g., HTAs), we searched the Cochrane library to find reviews on similar topics and examined the reviews to determine which review group undertook them. Two investigators undertook these classifications (AI, MS), with their results confirmed by a third reviewer (MA) with a clinical and research background.

Identification of new data for each review in the cohort. We performed systematic searches for each of the 100 reviews using a variety of electronic search strategies. Constructing searches as comprehensive as one would undertake for a formal systematic review (or an update) would involve a prohibitive amount of work given our cohort size of 100 systematic reviews.
Therefore, we adopted a combination of efficient strategies. Briefly, these involved developing simple subject searches and then limiting the results to the Core Clinical Journals subset plus the Randomized Controlled Trial publication type, subject searches run using the Clinical Query* filter in Ovid, applying the Related Articles function in PubMed® to the three largest and the three most recent studies in the original review (i.e., up to 6 studies in total), and using a ‘citing references’ search (through Scopus™) to identify new randomized trials that cited the original review. These search strategies served two purposes: one was to identify all new studies appropriate for updating the original systematic review; the other was to compare the performance of different strategies and evaluate their relative efficiency as surveillance methods for detecting signals for the need to update prior reviews. For studies where an updating signal occurred, we searched CENTRAL, The Cochrane Collaboration’s Central Register of Controlled Trials, using the subject search developed for MEDLINE®. Examples of subject searches with the limits tested are shown in Appendix C*. A sample recording sheet used as the basis for assessing search performance is shown in Appendix D.

For each systematic review in our cohort, project team members who had backgrounds in both medicine and research screened citations retrieved by the above methods to identify trials that would have met the inclusion criteria in the original meta-analysis. Retrieved records were screened in chronological order, and the full text of articles was used when necessary to determine eligibility or extract data. The review protocol stopped when one of the signals for the need for updating (defined below) was met. Wherever possible we identified new systematic reviews on the same topic. When the search strategies yielded no eligible new trials, we conducted more comprehensive electronic searches and reviewed relevant chapters in sources such as Clinical Evidence and UpToDate to ensure that we had not missed new sources of evidence. Figure 1 outlines the overall review protocol for assessing the presence or absence of signals for updating for each of the systematic reviews in the cohort.

**Outcomes: Signals for Changes in Evidence That Would Warrant Updating.** Ideally, assessments of the need to update previous systematic reviews would involve assessments by experts of new evidence relevant to the original review. Shekelle and colleagues used such an approach in order to determine if guidelines required updating.7 By choosing a small number of guidelines (17) produced by a single agency, they were able to ask the authors of the original guidelines to assess changes in evidence. This approach would clearly not be feasible for a larger sample (100 systematic reviews in the present case). It is also worth noting that identifying experts is not a straightforward task, requiring a balance of context expertise, methodological expertise, and freedom from bias regarding the question under consideration (not always easy to find among experts in a given area).

In designing a method for detecting changes in evidence without resorting to consulting experts, we considered the work of previous investigators12-14 who have addressed similar problems involving the comparison of two sets of results related to the question—randomized and non-randomized studies of the same intervention,14 initial and subsequent trials evaluating the same therapy,13 and conference proceedings versus full-length journal articles for the same

---

* In Ovid MEDLINE®, there are three clinical queries available for therapies; sensitivity, specificity and optimized. We used the optimized query.

trials. In all of these examples, investigators made determinations of important changes or differences between results without resorting to expert review. They achieved such determinations credibly by using a combination of quantitative signals (roughly the same as the ones we have chosen) and qualitative signals based on the language used to describe the results. For instance, if an article characterized a therapy as effective and another article evaluating the same therapy described it as ineffective, this would represent a major change. In a similar manner, we conceptualized quantitative and qualitative signals of potential changes in evidence sufficiently important to warrant updating of a previous systematic review.

Figure 1. Review protocol to detect signals for updating.

Search for new systematic reviews on same topic as cohort review

New systematic review identified?

Y

Search for new trials

Protocol searches

Supplemental searches

Screen candidate new trials and add eligible studies to meta-analysis

Quantitative signal detected? Qualitative signal from pivotal sources?

N

Candidate list of trials exhausted?

N

Consistent with commentaries, current guidelines, textbooks?

Y

Team conference

More searching needed?

Y

Finalize updating status

N

N
Quantitative signals consisted of changes in statistical significance (using the conventional alpha of 0.05) or large changes in effect size (a relative change in effect magnitude of at least 50%). We restricted these changes to those involving one of the primary outcomes of the original systematic review or any mortality outcome (i.e., all-cause mortality or any cause-specific mortality outcome for which the original review provided a meta-analytic estimate of effect). We also discounted ‘borderline’ changes in statistical significance, which we defined as having occurred when the original and updated meta-analytic results both had p-values in the range of 0.04 and 0.06. For instance, a change from $p = 0.041$ to $p = 0.059$ would not count as a quantitative signal to update, nor would the converse change (from $p = 0.059$ to $p = 0.041$). We discounted such changes, as well as changes in effect magnitude less than 50% and all changes involving non-primary outcomes, so that quantitative signals of changes in evidence would represent robust indicators of the need to update previous reviews. Quantitative signals were detected by performing updated meta-analyses that combined data from eligible new trials with the previous meta-analytic results.

Qualitative signals of the need to update involved factors relevant to the application of evidence beyond changes in the original meta-analytic estimates. These included new information about harm sufficient to impact clinical decision making, important caveats to the original results, emergence of a superior alternate therapy, and important changes in certainty or direction of effect. Qualitative signals were detected using explicit criteria for comparing the language used to characterize findings in the original systematic review with descriptions of findings in new systematic reviews that addressed the same topic, new ‘pivotal trials’, new clinical practice guidelines, or new editions of major textbooks (e.g., *UpToDate*). Pivotal trials were defined as trials that had a sample size at least three times the previous largest trial or were published in one of the 5 top general medical journals (New England Journal of Medicine, The Lancet, JAMA, Annals of Internal Medicine, and BMJ) based on a ranking by journal impact factor. We defined qualitative signals with two levels of importance: signals of ‘potentially invalidating changes in evidence’, which we considered as changes such that one would no longer want clinicians or policy makers to base decisions on the original systematic review (e.g., a pivotal trial characterizes treatment effectiveness in opposite terms to those in the original review); and signals of ‘major changes in evidence’, which we regarded as changes that would not completely invalidate the previous results but would still affect clinical decision making in important ways. Such changes might include information about the way the treatment must be delivered to confer benefit, identification of populations of patients for whom treatment is more or less beneficial, or information about impact on harder outcomes than those reported in the previous systematic review (e.g., the previous review analyzed intermediate endpoints, such as blood pressure or lipid levels, whereas new trials provide data on disease end-points, such as myocardial infarction or stroke, functional status, mortality).

Major changes also included changes in characterizations of effectiveness that were less extreme than those for potentially invalidating signals, but which would still affect clinical decision making. For example, whereas a change from ‘effective’ to ‘ineffective’ would represent a signal for a potentially invalidating change in evidence, a change from ‘possibly beneficial’ to ‘definitely beneficial’ would represent a major change. Importantly, no attempt was made to distinguish between varying descriptions of “possibly effective.” Characterizations such as “may be effective,” “promising,” “trends towards effectiveness,” and other similar phrases or concepts were all categorized as “possibly effective.” Thus, qualitative signals for changes in evidence
captured substantive differences in the characterization of treatment effects, not merely semantic differences.

Detailed definitions of the criteria for qualitative and quantitative signals are provided in Appendix A*; Appendix B provides specific examples of their application.

**Detection of Quantitative Signals for Updating**

An Excel worksheet was developed in which, for a given systematic review, project team members could enter the original meta-analytic result for each outcome into a template for the appropriate format (relative risk, odds ratio, risk difference, or weighted mean difference) and enter the results of new trials identified as eligible for inclusion. For a given outcome, the worksheet allowed entry of a summary estimate or raw data (e.g., a relative risk and 95% confidence interval or the number of events for patients in each study group using the format of a two by two table).

The worksheet was programmed to perform updated meta-analytic estimates and to apply logical tests to indicate when the updated result met one of the criteria for a quantitative signal (change in statistical significance or relative change in effect size of at least 50%). Because many of the original systematic reviews included a large number of trials and often did not report data for the individual trials in a complete fashion, it was impractical for us to obtain data for each trial included in the original meta-analytic estimate. Consequently, we performed the updated meta-analyses by combining the original pooled result with the individual results of eligible new trials. With fixed effects models for meta-analysis, this procedure gives the same result as would be obtained using the individual trials from the original meta-analysis Therefore, for pragmatic reasons (avoiding having to obtain original data from each trial included in each of the 100 systematic reviews) we employed fixed effects models in our updated meta-analyses. Though random effects models are usually preferred to avoid spurious precision in the face of heterogeneity, we regarded this approach as reasonable, since our goal consisted of detecting changes in evidence that had likely occurred, not producing exact estimates of updated treatment effects.

Data from new trials were entered into the meta-analytic calculator in chronological order, so that the time at which a quantitative signal was met could be identified. In general, we stopped the review protocol once a change in statistical significance or change in effect size of at least 50% occurred, though we sometimes continued to add new trials to confirm stability of the results.

**Group review and classification.** After assessment by the reviewer, each systematic review in the cohort was discussed at a case conference attended by the team of KS, MS, MA, and JJ. At this meeting, the final classification of the updating signal status was decided by group consensus, and the completeness of the evidence base was discussed. The team had the option to request additional searching, or search directly for new studies known or suspected by team members to be relevant.

**Date definitions for survival analysis.** Two survival analyses were undertaken. The first used the publication date as birth. We used the MEDLINE Entrez date as a surrogate for

publication date of the systematic review, as this date always includes a day, month, and year (not always the case for journal publication dates) and because the Entrez Date closely follows the publication date (typically within days to several weeks). In a second survival analysis, we defined birth as the end of the search period reported in the review. (This date did not always include a day and month. We imputed all missing months as June and all missing days as the 15th.) The end point, ‘death’ for both survival analyses was the Entrez date associated with the new evidence that resulted in the signal for updating. Where the updating signal derived from non-MEDLINE sources (e.g., an advisory from the Centers for Disease Control and Prevention, the Food and Drug Administration, or a chapter in a textbook), we used the date of publication as the date of the signal for updating. For surviving systematic reviews, observations were censored on September 1, 2006, the approximate midpoint of the 4-month period during which searches were performed for the entire cohort.

**Performance of the surveillance searches in detecting signaling evidence.** Three main types of signaling evidence were used: new RCTs that were added to a meta-analysis from the original systematic review in the manner of a cumulative meta-analysis, single RCTs that met our criteria for a pivotal trial, and new systematic reviews that provided evidence that appeared to overturn the findings of the original review, either by contradicting the original findings, adding an important caveat or demonstrating a significant harm. Other signaling evidence (i.e., evidence that provided the basis for signals) included FDA advisories and expert opinion from *UpToDate*, and clinical trials that did not meet the criteria for pivotal trial. These sources were used as sources of signals for updating in only five reviews.

The surveillance searches looked for primary studies and for systematic reviews with the publication type meta-analysis in MEDLINE. To determine the effectiveness of these searches to detect signaling evidence, we examined recall of signaling articles in the subset of systematic reviews studies here were those updated by search (n=79), and for which a major or notable signal occurred. For the analysis of RCTs added to the cumulative meta-analysis, only those systematic reviews which also had a quantitative signal and were updated by search were studied.

Any signaling evidence added by nomination was tested to determine if it was indexed in MEDLINE and if would have been retrieved by the searches. In some cases, the evidence was published after the searches for new evidence for that systematic review were run. The database was updated with the search results for those nominated publications.

Targets for the cumulative meta-analysis were any RCT added to the meta-analysis of the outcome which had the signal, up to the point where the signal occurred. Targets for the final RCTs were the pivotal RCTs. Targets for final MAs were the newer meta-analyses that contained the evidence that rendered the cohort systematic review potentially in need of update. These were meta-analyses that were not explicit updates. Finally, all signaling evidence was considered. For each of these analyses, recall was calculated for each type of search. For the final analysis of recall of any signaling evidence, two additional variables were created representing recall from MEDLINE by any of the subject search methods (CQ, AIM RCT or MA) and recall from MEDLINE by either of the related articles search methods (RI RCT and RI MA).
II. Publication Time Lags

For question 3, the impact of publication time lags on updating, we supplemented the data set used in the survival analysis with additional eligible systematic reviews identified through *ACP Journal Club*, as well as AHRQ Evidence Reports that met all eligibility criteria for the main cohort, except inclusion in *ACP Journal Club*.

We determined dates for performance of the original search, manuscript acceptance, and publication of the review. We regarded the search date as the most recent date reported in the methods section of the systematic review. For Cochrane reviews, we used the most recent of the following dates: the search date reported in the search strategy section in the body of the review, the date new studies were found and included/excluded (e.g., for updated reviews), or the date new studies were sought but not found. For database dates, the end date reported for MEDLINE searching was used (i.e., 1966-June Week 4, 2003) if available. If the MEDLINE date was not reported, any other database end date was used. If no end date was reported, the variable was treated as missing. For all types of reviews, the publication date and indexing date was taken from the Ovid MEDLINE records.

For each date (original search, manuscript acceptance, publication), we identified a year, month, and day. When month was missing, we imputed the 6th month; when day was missing, we imputed the 15th day of the month.

III. Growth of the Literature by Clinical Area

MEDLINE searches based on high-level MeSH headings corresponding to the ISI journal categories were undertaken. The resulting set of citations was limited to the publication type Randomized Controlled Trial, to the publication type Clinical Trial but not Randomized Controlled Trial, to the MEDLINE Systematic Review subset, and to the publication type Clinical Practice Guidelines. Searches were then limited by year for each year between 1988 and 2006. We chose 1988 as the beginning of the time period of interest, as this date corresponded to the period five years prior to the earliest search date for any systematic review in the cohort. Search strategies are illustrated in Appendix E*.

IV. Survey of Organizations Engaged in Funding or Production of Systematic Reviews

This exploratory Internet pilot survey on current updating practices and policies employed a purposeful sample to allow for investigation of likely information-rich cases. We chose 9 organizations well known to fund or carry out systematic reviews, as well as 12 EPCs, in addition to AHRQ, were also asked to complete this survey. The identities of the organizations have been kept anonymous per the statements contained in the informed consent signed by participants in the survey and as stipulated in the research protocol approved by the institutional ethics review board at the Children’s Hospital of Eastern Ontario.

The survey was provided to participants via the Survey Monkey\textsuperscript{15} web-based service. This was considered a suitable forum given distribution of the sample across a wide international geographical area, and that key informants are frequent Internet users with email addresses.\textsuperscript{16,17} Emails were sent directly to organizational Directors or to the highest ranking scientific or administrative official, asking them to identify the most appropriate internal respondent to answer the questionnaire. Data collection consisted of approximately 50 questions (including skip-logic functionality). These questions focused on the following topics: (a) updating policies, (b) responsibility for updating, (c) estimates of outdated reviews, (d) updating strategies and practices, including when to update, surveillance and triggers impacting updating decisions, (e) strategies for how to conduct an update, (f) barriers and facilitators to this process, (g) views on updating collaboration between groups and (h) descriptive demographics and characteristics of the organization and the representative key informant. It was estimated the survey took between 20 to 30 minutes to complete. (Appendix G\textsuperscript{*}: Survey Instrument)

We attempted to increase our overall response rate by employing recommended survey methods to maximize Internet survey participation.\textsuperscript{17-20} Participants were contacted four times. A small financial incentive was offered to all participants who completed the survey. On clicking on the link to the survey, participants were presented with a description of the purpose of the study, assurance of confidentiality, and a statement of the research protocol by the hospital ethics review board, followed by a request to provide informed consent or decline participation in the survey. Reminder emails were scheduled for day 10, 15 and 25 of the survey.

Analysis

We fit non-parametric Kaplan-Meier curves to the data set of censored and uncensored observations and used multivariable proportional-hazards models to examine the association between survival and various features of the systematic reviews at the time of publication. We distinguished two categories of potential predictors of survival. The first category consisted of features knowable at the time of publication for a given systematic review, including clinical content area (e.g., cardiovascular medicine, obstetrics and gynecology, critical care, infectious diseases), numbers of participants and trials included in the meta-analysis, the identification of heterogeneity or publication bias, ‘recent or ongoing activity in the field’, which we defined as present if the review included at least one trial published within the last year of its search period or if the review identified ongoing trials eligible for inclusion. Because some evidence exists to suggest that Cochrane reviews differ in important ways from other systematic reviews,\textsuperscript{1} we also included a dichotomous variable for Cochrane review versus other systematic reviews. The second category of predictors consisted of features knowable only after some surveillance of the literature (but not performance of a full update of the review). Such predictors included the number of new trials eligible for inclusion in an update of the original review, the number of new participants in these trials, the ratio of the new total number of trials to the previous total, and the ratio of the new total number of participants to the previous total.

After confirming that the assumption of proportionality applied, we performed stepwise multivariate analyses using a threshold of $p \leq 0.1$ for variable selection and retention. In addition to the proportional hazards analysis to estimate predictors of survival, we conducted logistic

regression analysis to identify predictors of survival less than two years. Cohort members that were censored in less than two years were counted as missing for this analysis. All analyses were performed with SAS version 9.0 (The SAS Institute, Cary, North Carolina).

Analysis of group differences in time lags in the publication process was made using nonparametric statistics (e.g., Kruskal-Wallis test for differences in median publication times between groups).

**Survey Analysis**

Closed-ended questions were analyzed primarily using a descriptive summary of findings in the form of frequencies. In addition, percentages were calculated and other details reported in text and tabular form. Participating organizations were not identified in the results as only aggregate data is reported. The EPCs also responded to several additional open-ended questions of particular interest to the AHRQ’s EPC Program. The responses to these supplemental questions were compiled for internal use by the AHRQ and are therefore not discussed in this report.
Chapter 3. Results

Results of Literature Search and Cohort Screening

Records for 651 potentially eligible systematic reviews were identified through searching. Achieving our target sample size of 100 reviews for the analysis of updating signals required that we assess a total of 325 reviews for eligibility. We screened additional reviews to add a further 50 reviews to the set of reviews in the analysis of publication time lags. (The analysis of time lags was less labor intensive, permitting a larger cohort size for this part of the project). 165 records were excluded on the basis of the ACP Journal Club record, and 60 articles were excluded after assessment of the full article. Exclusion reasons are shown in Figure 2.
**Figure 2. Flow of information through eligibility assessment.**

*This category includes reviews not focused on a specific class of drug, device or procedure, as well as ones focused on educational or behavioral interventions, or complementary therapies.

†This category includes updates of systematic reviews already in cohort, topics similar to that of a systematic review already included, or the journal version of an included Cochrane review

‡ This category includes meta-analysis using individual patient data without regular meta-analysis, meta-regression, or indirect meta-analysis.
Assessment of the New Evidence

Seventy-seven of the systematic reviews were assessed against new evidence found (if any) through searching and 23 were assessed against an updated systematic review. The updated review could be either an update performed by the authors of the original review, or a newer review on the same topic identified through the search of ACP Journal Club that would itself be eligible for inclusion in the cohort.

Characteristics of Included Studies

Composition of the Cohort

Each review in the cohort of 100 systematic reviews included a median of 13 studies (inter-quartile range: 8 to 21) and 2663 participants (inter-quartile range: 1281 to 8371). We were able to identify at least one new eligible trial for 85 systematic reviews, with a median of 4 new trials (inter-quartile range: 1 to 7) and 1160 new participants (inter-quartile range: 170 to 3689) per review. The five most common clinical content areas for the original systematic reviews were cardiovascular medicine (20), gastroenterology (13), neurology (11), infectious diseases (9), and respiratory system (9). Only 15 of the reviews evaluated the effects of medical devices or procedures; drug therapies provided the focus for the rest of the cohort (Table 1).
Table 1. Characteristics of the Cohort of 100 Systematic Reviews

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Composition of Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication Type</td>
<td>Peer-reviewed journal article (72), Cochrane review (27), Other* (1)</td>
</tr>
<tr>
<td>Intervention Type</td>
<td>Medication (85), Medical device (8), Procedure (7)</td>
</tr>
<tr>
<td>Clinical Categories</td>
<td>Cardiovascular (20), Gastroenterology (13), Neurology (11); Other 10 categories each included fewer than 10 systematic reviews</td>
</tr>
<tr>
<td>Publication Period</td>
<td></td>
</tr>
<tr>
<td>January 1995 to February 28, 1997</td>
<td>16</td>
</tr>
<tr>
<td>March 1997 to April 30, 1999</td>
<td>22</td>
</tr>
<tr>
<td>May 1999 to June 30, 2001</td>
<td>25</td>
</tr>
<tr>
<td>July 2001 to September 30, 2003</td>
<td>21</td>
</tr>
<tr>
<td>October 2003 to December 31, 2005</td>
<td>16</td>
</tr>
<tr>
<td>Dates of searches</td>
<td></td>
</tr>
<tr>
<td>June 1990 to April 1993</td>
<td>2</td>
</tr>
<tr>
<td>May 1993 to March 1996</td>
<td>21</td>
</tr>
<tr>
<td>April 1996 to February 1999</td>
<td>28</td>
</tr>
<tr>
<td>March 1999 to January 2001</td>
<td>30</td>
</tr>
<tr>
<td>February 2002 to December 2004</td>
<td>18</td>
</tr>
<tr>
<td>Search date not reported</td>
<td>1</td>
</tr>
<tr>
<td>Source</td>
<td>The Cochrane Library (27), BMJ (19), JAMA (7), Lancet (7), Annals of Internal Medicine (6), Archives of Internal Medicine (5) and 25 other titles</td>
</tr>
<tr>
<td>Median Number Included Trials</td>
<td>13 (inter-quartile range: 8-21)</td>
</tr>
<tr>
<td>Included Number Included Participants</td>
<td>2663 (inter-quartile: 1281-8371)</td>
</tr>
<tr>
<td>Included ≥ 1 trial published within last year of search period</td>
<td>67 reviews</td>
</tr>
<tr>
<td>Original review mentioned ongoing trials</td>
<td>26 reviews</td>
</tr>
</tbody>
</table>
| Heterogeneity as assessed by authors of original review | Identified as statistically significant: 50  
Not statistically significant but authors still suspected as possibly present: 11  
Heterogeneity regarded by authors as absent: 32  
Not assessed: 7                                                                 |
| Publication bias as assessed by authors of original review | Identified as statistically significant: 4  
Not statistically significant but authors still suspected as possibly present: 14  
Regarded by authors as absent: 22  
Not assessed: 60                                                                 |

* Other: 1 systematic review published by the Canadian Coordinating Office for Health Technology Assessment
Signals for Updating

Of the 100 systematic reviews, a quantitative signal for updating occurred in 30 cases. Qualitative signals for the need to update occurred in 54 cases, including 8 that met criteria for a potentially invalidating change in evidence and 46 that met criteria for a major change. Qualitative signals had their basis in new systematic reviews in 23 cases (including explicit updates in 5), pivotal trials in 25 cases, and other sources in 6 cases (trials discussed in ACP Journal Club or UpToDate and advisory statements issued by the Centers for Disease Control and Prevention or the Food and Drug Administration). The primary event of interest, a quantitative signal involving the primary outcome of the original systematic review or qualitative...
signal for potentially invalidating or major changes in evidence, occurred for 57 reviews (57%; 95% CI: 47% to 67%) in the cohort (Table 2).

Table 2. Frequency of the Different Types of Signals for Updating

<table>
<thead>
<tr>
<th>Type of Signal for Updating</th>
<th>Number of systematic reviews in cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative signal</td>
<td>20</td>
</tr>
<tr>
<td>Change in statistical significance</td>
<td>18</td>
</tr>
<tr>
<td>Relative change in effect size ≥ 50%</td>
<td>12</td>
</tr>
<tr>
<td>Qualitative signal</td>
<td>54</td>
</tr>
<tr>
<td>Opposing findings</td>
<td>7</td>
</tr>
<tr>
<td>Substantive changes short of opposition</td>
<td>16</td>
</tr>
<tr>
<td>Clinically important caveats</td>
<td>28</td>
</tr>
<tr>
<td>Clinically important expansion of therapy</td>
<td>3</td>
</tr>
<tr>
<td>Harm that completely undermines therapy</td>
<td>1</td>
</tr>
<tr>
<td>Superior alternate therapy</td>
<td>1</td>
</tr>
<tr>
<td>Primary event of interest: either Qualitative or Quantitative signal</td>
<td>57</td>
</tr>
</tbody>
</table>

Survival Analysis

Using publication date as ‘birth’, median event-free survival (i.e., time without a signal for updating) was 5.5 years (95% CI: 4.6-7.6). However, in 23 cases, signals for updating occurred in less than 2 years, and in 15 cases the signal occurred in less than 1 year. In 7 cases, a signal had already occurred at the time of publication of the original systematic review (in one case, 295 days prior to publication).
Figure 4. Kaplan Meier plot showing the overall event free survival (time without a signal for updating) using publication date as ‘birth’; the immediate drop in survival at time zero reflects the 7 systematic reviews for which signals for updating had already occurred at the time of publication. Symbols represent censored cases.

In univariate analyses, shorter survival was associated with a clinical content area of cardiovascular medicine (hazard ratio [HR] of 2.58, 95% CI: 1.39 to 4.78; \( p = 0.003 \)), increase in the total number of patients by a factor of 2 or more (HR of 1.79; 95% CI: 1.03 to 3.10; \( p = 0.04 \)), and heterogeneity for at least one outcome in the original systematic review (HR of 1.64, 95% CI: 0.94 to 2.86; \( p = 0.08 \)) (Figures 4-7). Other potential predictors evaluated, but not found to significantly affect survival included: the number of included patients in the original review, identification of publication bias in the original review, the inclusion of at least one trial published in final 12 months of the search period, the identification of ongoing trials, and the publication type (Cochrane reviews versus those published in peer review journal articles) (Table 3).
Table 3. Univariate Survival Analysis

<table>
<thead>
<tr>
<th>Clinical Category</th>
<th>Hazard Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurology</td>
<td>1.37 (0.59, 3.16)</td>
<td>0.47</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>2.58 (1.39, 4.78)</td>
<td>0.003</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>1.35 (0.58, 3.13)</td>
<td>0.48</td>
</tr>
<tr>
<td>Other</td>
<td>reference</td>
<td>-</td>
</tr>
<tr>
<td>Heterogeneity present or suspected</td>
<td>1.64 (0.94, 2.86)</td>
<td>0.08</td>
</tr>
<tr>
<td>Publication bias present or suspected</td>
<td>0.99 (0.46, 2.12)</td>
<td>0.98</td>
</tr>
<tr>
<td>Activity in field *</td>
<td>1.36 (0.76, 2.44)</td>
<td>0.30</td>
</tr>
<tr>
<td>Number of included studies &gt; median (13) for cohort</td>
<td>0.79 (0.46, 1.33)</td>
<td>0.37</td>
</tr>
<tr>
<td>Number of included participants &gt; median (2663) for cohort</td>
<td>1.22 (0.72, 2.06)</td>
<td>0.47</td>
</tr>
<tr>
<td>Ratio of New to Original Total N &gt; 2†</td>
<td>1.79 (1.03, 3.1)</td>
<td>0.04</td>
</tr>
<tr>
<td>Any of 3 criteria for substantial increases in number of new trials or patients‡</td>
<td>1.17 (0.59, 2.32)</td>
<td>0.66</td>
</tr>
<tr>
<td>Cochrane review</td>
<td>0.74 (0.39, 1.39)</td>
<td>0.35</td>
</tr>
<tr>
<td>Largest new trial larger than previous largest N trial</td>
<td>1.04 (0.99, 1.1)</td>
<td>0.09</td>
</tr>
</tbody>
</table>

*Recent activity defined as present if original systematic review included at least 1 trial published within the final 12 months of the search period or if original systematic review identified ongoing trials eligible for inclusion. This variable was coded as present for 71 of the included systematic reviews.

† Ratio of New to Original Total N > 2 (i.e., increase in total sample size by more than a factor of 2)

‡ Size criteria C1-C3 defined as any of the following occurring: increase in total number of trials by $\geq 50\%$, increase in total number of participants by $\geq 50\%$, publication of a new trial with sample size $\geq 3$ times size of previous largest trial.

Table 4 shows the results of multivariate analysis with adjustment for all variables shown. We also performed stepwise multivariate analysis using a threshold of $p \leq 0.1$ for variable selection and retention, which resulted in a model in which the following variables predicted decreased survival: clinical content area of cardiovascular medicine (HR of 3.26, 95% CI: 1.71 to 6.21; $p = 0.0003$), heterogeneity in the original systematic review (HR of 2.23, 95% CI: 1.22 to 4.09; $p = 0.01$), and the ratio of the largest new trial to the largest trial from the original review (HR of 1.08, 95% CI: 1.02 to 1.15; $p = 0.01$). Systematic reviews with more than the median of 13 included studies had increased survival (HR of 0.55; 95% CI: 0.31 to 0.98; $p = 0.04$).

In logistic regression analysis no variable significantly affected the risk of a signal for updating occurring within 2 years of publication, though trends towards increased risk were observed for cardiovascular topics (odds ratio of 2.67; 95% CI: 0.88 to 8.1, $p = 0.08$) and an increase in the total number of patients by at least factor of 2 (odds ratio of 2.29; 95% CI: 0.84 to 6.25, $p = 0.11$). A trend towards decreased risk of a signal for updating occurring within 2 years was seen for systematic reviews with more than the median of 13 included studies (odds ratio of 0.38; 95% CI: 0.14 to 1.04; $p = 0.06$). Varying the time period of interest (e.g., predicting a signal for updating within 1 year or 3 years of publication) did not substantially alter the results.
<table>
<thead>
<tr>
<th>Clinical Category</th>
<th>Hazard Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurology</td>
<td>1.38 (0.52, 3.70)</td>
<td>0.52</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>3.09 (1.47, 6.52)</td>
<td>0.003</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>1.44 (0.57, 3.62)</td>
<td>0.44</td>
</tr>
<tr>
<td>Other</td>
<td>reference</td>
<td>-</td>
</tr>
<tr>
<td>Heterogeneity present or suspected</td>
<td>2.22 (1.21, 4.08)</td>
<td>0.01</td>
</tr>
<tr>
<td>Publication bias present or suspected</td>
<td>1.06 (0.47, 2.41)</td>
<td>0.89</td>
</tr>
<tr>
<td>Activity in field *</td>
<td>1.45 (0.74, 2.82)</td>
<td>0.28</td>
</tr>
<tr>
<td>Number of included studies &gt; median (13) for cohort</td>
<td>0.42 (0.21, 0.81)</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of included participants &gt; median (2663) for cohort</td>
<td>1.56 (0.79, 3.08)</td>
<td>0.20</td>
</tr>
<tr>
<td>Ratio of New to Original Total N &gt; 2†</td>
<td>1.86 (0.95, 3.61)</td>
<td>0.07</td>
</tr>
<tr>
<td>Any of 3 criteria for substantial increases in number of new trials or patients‡</td>
<td>0.97 (0.45, 2.12)</td>
<td>0.94</td>
</tr>
<tr>
<td>Cochrane review</td>
<td>1.35 (0.62, 2.97)</td>
<td>0.45</td>
</tr>
<tr>
<td>Largest new trial larger than previous largest N trial</td>
<td>1.06 (0.99, 1.14)</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*Recent activity defined as present if original systematic review included at least 1 trial published within the final 12 months of the search period or if original systematic review identified ongoing trials eligible for inclusion.

† Ratio of New to Original Total N > 2 (i.e., increase in total sample size by more than a factor of 2)

‡ Size criteria C1-C3 defined as any of the following occurring: increase in total number of trials by ≥ 50%, increase in total number of participants by ≥ 50%, publication of a new trial with sample size ≥ 3 times size of previous largest trial

Survival contrasting the significant predictors with the rest of the cohort is illustrated in Figures 5 through 7.
Figure 5. Kaplan Meier plot showing survival by clinical topic area of the original systematic review, stratified by cardiovascular (n=20 reviews) versus all other topics (n=80). Symbols represent censored cases.
Figure 6. Kaplan Meier plot showing survival stratified by the presence or absence of heterogeneity in the systematic review; statistical heterogeneity was identified as definitely or likely present for at least one outcome in 61 of the 100 reviews. Symbols represent censored cases.

Median Survival (95% CI)
- Not Heterogeneous: 7.6 (5.1, --)
- Heterogeneous: 4.9 (3.5, 6.6)
Figure 7. Kaplan Meier plot showing the effect on survival of a doubling of the total number of patients (i.e., ratio of new total sample size to old total > 2), which occurred for 25% of systematic reviews in the cohort. Symbols represent censored cases.

When survival analyses were repeated using the end of the search period as ‘birth’, rather than the publication date, the median survival was 6.9 years (95% CI: 6.1 to 9.0), with a median time to a signal for updating of 4.3 years (inter-quartile range: 2.1- 6.4 years). The signal for updating occurred within 1 year of the search in 4 cases, within 2 years of the search in 11 cases and within 3 years of the search in 20. Predictors of increased or decreased survival did not differ from those identified in the analysis that used publication date as ‘birth.’

**Directions of Changes in Evidence and Expected Impact on Practice**

Of the 18 reviews with changes in statistical significance, 13 involved a gain of statistical significance (i.e., a previously non-significant result became statistically significant) and 5 involved a loss of significance. For the 12 reviews with a relative change in effect size of at least 50%, 3 involved an increase in effect magnitude and 9 involved a decrease in effect. However,
because these outcomes could involve harms or benefits, we also characterized the expected impact on practice of the changes that gave rise to the signal for updating. Increases in magnitude of benefit, certainty about benefit, or identification of new patient populations that benefit from the treatment were classified as expected increases in therapeutic application. Decreases in magnitude of benefit, decreased certainty about benefit, findings of increased harm or other limitations on benefit were all classified as leading to decreased therapeutic application. Using such explicit criteria, use of the therapies evaluated would be expected to increase in 19 and decrease in 28 (Table 5).

We also assessed the impacts on certainty of results due to the changes in evidence that gave rise to signals for updating. We characterized changes in certainty using the 5-point scale that formed the basis for judging characterizations of effectiveness (Appendix A*). This scale included the following categories: definitely effective, probably or possibly effective, uncertain effectiveness, probably or possibly ineffective, and definitely ineffective. When the updated result lay further from the middle position (complete uncertainty) than the original result, we regarded certainty as having increased. Conversely, when the updated result lay closer to the middle position than the original result, we regarded certainty as having decreased. When the updated and original results were equally distant from the middle position (e.g., definitely effective and definitely ineffective), we did not regard certainty as having changed. Such cases would, however, count as impacting therapeutic use. As shown in Table 5, the majority of signals for updating involved increases in certainty (30 reviews) or no changes in certainty (50 reviews).

### Table 5. Changes in Certainty and Expected Impacts on Practice Associated with Signals for Updating.

<table>
<thead>
<tr>
<th>Impacts of new evidence on certainty of results</th>
<th>Expected impacts of new evidence on clinical practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in certainty</td>
<td>Increased therapeutic use</td>
</tr>
<tr>
<td>Decrease in certainty</td>
<td>Decreased therapeutic use</td>
</tr>
<tr>
<td>Unchanged certainty</td>
<td>Unchanged therapeutic use</td>
</tr>
<tr>
<td>Unclear change</td>
<td>Unclear change</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impact</th>
<th>Count</th>
<th>Expected Impact</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in certainty</td>
<td>30</td>
<td>Increased therapeutic use</td>
<td>19</td>
</tr>
<tr>
<td>Decrease in certainty</td>
<td>3</td>
<td>Decreased therapeutic use</td>
<td>28</td>
</tr>
<tr>
<td>Unchanged certainty</td>
<td>50</td>
<td>Unchanged therapeutic use</td>
<td>44</td>
</tr>
<tr>
<td>Unclear change</td>
<td>17</td>
<td>Unclear change</td>
<td>9</td>
</tr>
</tbody>
</table>

### Search Performance

Across all reviews, 477 new reports were identified as eligible for inclusion to the systematic reviews. Of these, searching identified 430, and 47 were identified by the reviewers from among the studies included in meta-analysis retrieved by the subject search or related item searches. Forty of these nominations (85%) were indexed in MEDLINE, thus the searches retrieved 92% of eligible new studies identified. Forty-three of the 47 missed studies were from systematic reviews where we had searched CENTRAL. Two of these 43 nominations were retrieved by the CENTRAL search.

In 59 cases, a single search strategy would have been sufficient to retrieve all eligible new studies found by any method. Related article RCTs was sufficient in 45 cases, Clinical Query in 34, Core Clinical Journal RCTs in 9, CENTRAL in 14 and Citing RCTs in 3. Systematic reviews with multiple sufficient strategies tended to be those with few new studies. In 68 cases, searching Related Article RCT and Clinical Query would have retrieved all studies either because one strategy or the other was sufficient or because the two together was sufficient.

The median number of records retrieved by the combination of Related Article RCT and Clinical Query by the date at which the signal for updating was detected was 71 (1st and 3rd quartiles; 25, 106). The median number of records retrieved by this combination and assessed as on topic was 7 (1st and 3rd quartiles; 4, 24).

To identify newer quantitative systematic reviews, the Related Article search and the subject search were limited to publication type of meta-analysis. The Related Article search recalled 45% of the meta-analyses found to be on topic and the subject search limited to the meta-analysis publication type identified 66% of the on topic meta-analyses. Of the records assessed by the review team, precision (proportion of assessed records found to be relevant) of the subject search was 0.38 and precision of the Related Article MA search was 0.36.

**Performance of the surveillance searches in detecting signaling evidence.** There were 27 final RCTs in cohort systematic reviews that were updated by searching and had a qualitative signal of major or notable. Sixteen of these also had a quantitative signal and so formed the basis of the analysis of success in detecting RCTs added to the cumulative meta-analysis.

Six of the 27 final RCTs were by nomination and the remaining 21 were found by the search. Three of the nominations were recent, high profile trials. These were used rather than reviewing the candidate list, thus for the purpose of evaluating search performance. As these occurred after the search date, these three were tested to see if the search would have retrieved them, and whether they cited the cohort systematic review. The remaining 3 nominated final RCTs were identified through meta-analysis. There were 34 targets for studies added to cumulative meta-analysis; 27 were candidates found through searching, 5 were nominations, 2 were meta-analyses found through searching where the individual trial data could not be extracted. One of the nominations was a trial published after the search date, and was manually tested to see which searches would have retrieved it. There were 9 signaling meta-analyses – one was nominated, all others were identified through searching.

Other signaling evidence was used in only 5 reviews that were updated by search. Evidence included FDA advisories and expert opinion from *UpToDate*, and clinical trials that did not meet the criteria for pivotal trial. Three of these 5 sources were indexed in MEDLINE. Two were identified through searching.

Recall by each search of each type of evidence is shown in Table 6. Retrieval was best for RCT and MA evidence, but the search methods did retrieve some of the other evidence. Overall search performance of final evidence stood at 0.65 for subject search methods, 0.76 for related article methods, 0.55 for CENTRAL and 0.17 for citing references. Across all applications in which the citing reference technique was tested, its strongest performance was in detecting other final evidence, with 0.33 recall. One of the highest recall scores seen in this study was recall of 0.89 for final RCTs found through related article searching. In general, search methods showed somewhat higher recall for final evidence than for all evidence found relevant to the reviews, and the relative performance of the various methods was similar to that seen in the more general context.
Most information was found through searches. Of 62 pieces contributing to the signal, 57 (92%) were identified through the searches of MEDLINE. The additional material was an included study in a systematic review identified through searching, known to the team, or UpToDate.

<table>
<thead>
<tr>
<th>Table 6. Recall of Signaling Evidence by the Surveillance Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In Quantitative Signal n=34</strong></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Related Article search with RCT limit</td>
</tr>
<tr>
<td>Subject search limited to Core Clinical journals and RCT publication type</td>
</tr>
<tr>
<td>Citing Reference search</td>
</tr>
<tr>
<td>Subject search with Clinical Query limit</td>
</tr>
<tr>
<td>Subject search with meta-analysis publication type limit</td>
</tr>
<tr>
<td>Related Article search with meta-analysis limit</td>
</tr>
<tr>
<td>Retrieved by any subject search method</td>
</tr>
<tr>
<td>Retrieved by any related article method</td>
</tr>
<tr>
<td>Indexed in MEDLINE</td>
</tr>
<tr>
<td>Found by any search</td>
</tr>
</tbody>
</table>

**Adequacy of MEDLINE coverage for surveillance.** While there is general agreement that searching a single database is inadequate for developing the evidence base for systematic reviews, the adequacy of MEDLINE for detecting the need to update (surveillance searching) has not been previously examined. We consider the proportion of studies in the original reviews that were indexed in MEDLINE, the survival of those in known updates from this sample, and the proportion of new relevant studies identified from any source that were indexed in MEDLINE.

**Original systematic reviews:** Of 2065 reports included in the original systematic reviews, 407 (25%) were not indexed in MEDLINE. MEDLINE indexed publications accounted for 89% of total number of participants (N) included in the original systematic reviews, although we could not identify values for N in all cases, and 40% of cases with missing N were for non-indexed studies. For reports where we could identify N, the median size was larger for MEDLINE indexed studies compared with non-indexed studies (116 participants [inter-quartile range: 43-365] vs. 80 participants [40-224]).
**Updated Cochrane reviews.** Of original Cochrane reviews assessed through an update, 95 studies were indexed in MEDLINE. Of these 95 studies, 4 (4%) were excluded by the author in the update. Among the 56 studies not indexed in MEDLINE, 13 (23%) were excluded in the update (odds ratio 0.145, CI 0.045-0.472), suggesting that material from sources not indexed in MEDLINE may become less important over time.

**New studies.** The indexing status and number of new studies assessed as eligible for inclusion in the reviews were considered. New studies included candidates identified through searching, nominations found through newer meta-analyses or known to our team, and studies included in explicit updates. Of 590 studies assessed as eligible, 33 (6%) were not indexed in MEDLINE. These reports accounted for 5503 of 648531 new participants (N) identified (1%). All pivotal trials, those RCTs that, by themselves, provided in signal for update, were indexed in MEDLINE (n=19).

### Time Lags in the Production and Publication of Systematic Reviews

One hundred and forty-eight reports were included in this analysis, of which 91 (62%) were journal published reviews, 36 (24%) were Cochrane reviews and 21 (14%) were HTA reports. Of HTA reports, 19 (90%) were AHRQ evidence reports. For Cochrane reviews, we used the most recently published version of the Cochrane review.

The median time from last reported search date to indexing was 75 weeks with an inter-quartile range of 52 to 111 weeks. Lag from last search date to publication is shortest for Cochrane reviews (median 31 weeks, inter-quartile range: 22-65) and longest for journal reviews (median 69 weeks; inter-quartile range: 44-92), with technical reports falling in between (median 58 weeks; inter-quartile range: 45-74) (Kruskal Wallis $\chi^2$ 11.24, p =0.004) For reviews assessed for need of update, 7 were found to have gone out of date by the time of publication.

Intermediate milestones of submission and acceptance dates were reported only for journal published reviews, but reveal what proportion of total preparation time is under the control of investigators. For journal-published reviews where submission and publication dates are known (n=17) median processing time was 41 weeks (inter-quartile range; 29-55 weeks) weeks and where acceptance and publication dates are known (n=55) median processing time was 18 weeks (inter-quartile range; 13-27 weeks). The difference gives some indication of the time taken in peer review.

The 3 journal-published and 6 Cochrane reviews that reported more than one search date showed shorter lags from last search date to publication than those that did not appear to have updated the search. Eight HTA reviews reported updating their search and 11 did not, but the lags from most recent search to publication were essentially the same. Still, there was a significant overall effect by level of search updating (Log Rank (Mantel-Cox), Chi-square 7.253, df=1, p =0.007).

Publication lags were assessed in the main cohort to examine and trends over time. There was an apparent trend towards decreased publication lags over time, with more recent publication dates having shorter publication lags (p =0.12). However, this reflected bias sampling in the sense that the only way for a recent article to be sampled for inclusion in the cohort would be by having a short publication lag. In other words, systematic reviews initiated in, say, 2004, could only end up in the cohort, if they had relatively short delays before
publication. To avoid this bias, we analyzed the relationship between publication date and publication lag using only systematic reviews published prior to January 1, 2003. In this analysis, the trend towards shorter publication lags with more recent reviews disappeared completely, with a much smaller regression coefficient and p-value > 0.8.

**Publication Velocity**

The patterns of evidence accumulation at the macro level (by clinical area), or at the micro level (within a particular systematic review) could help to identify or predict optimal update intervals. Velocity at the macro level is considered here.

![Figure 8. Growth of controlled trials, RCTs, systematic reviews and clinical practice guidelines, 1988-2006.](image)

The three clinical areas with greatest representation in the cohort are cardiac and cardiovascular disease, neurology and gastroenterology (Table 1). Growth of randomized controlled trials, other controlled trials, appear linear in this time frame (Figure 8, Table 6).

Publication doubling times, when calculated under the assumption of linearity, were consistent across clinical content areas and increased markedly as the time from the series start increased (Table 7). All series shown here begin in 1988. For example, approximately 981 new oncology trials were published in 1988. This number doubled in a little over two years (2.2) in 1990 and will take almost 20 years (18.4) for those studies published in 2005.
**Table 7. Linear Fit And Rate Of Growth By Clinical Area, Doubling Time In Years From Various Starting Years Assuming Linear Growth.**

<table>
<thead>
<tr>
<th>Clinical area</th>
<th>Pearson R Cumulative RCTs &amp; CCTs</th>
<th>Rate of new RCTs per year</th>
<th>Doubling time from starting year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncology</td>
<td>1.00</td>
<td>981</td>
<td>2.2 6.0 11.1 18.4</td>
</tr>
<tr>
<td>Clinical Neurology</td>
<td>0.98</td>
<td>1158</td>
<td>1.9 4.4 9.6 18.4</td>
</tr>
<tr>
<td>Peripheral Vascular Disease</td>
<td>1.00</td>
<td>1238</td>
<td>2.3 6.4 11.5 18.2</td>
</tr>
<tr>
<td>Infectious Diseases</td>
<td>1.00</td>
<td>1227</td>
<td>2.0 6.0 11.4 17.6</td>
</tr>
<tr>
<td>Respiratory System</td>
<td>1.00</td>
<td>866</td>
<td>2.0 5.9 11.4 17.8</td>
</tr>
<tr>
<td>Cardiac and Cardiovascular Systems</td>
<td>0.99</td>
<td>814</td>
<td>2.1 5.3 10.6 18.3</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>0.99</td>
<td>784</td>
<td>1.9 4.7 9.4 18.7</td>
</tr>
<tr>
<td>Endocrinology and Metabolism</td>
<td>0.99</td>
<td>581</td>
<td>2.0 3.9 8.8 19.0</td>
</tr>
<tr>
<td>Gastroenterology and Hepatology</td>
<td>1.00</td>
<td>515</td>
<td>2.4 8.0 11.5 18.0</td>
</tr>
<tr>
<td>Obstetrics and Gynecology</td>
<td>1.00</td>
<td>400</td>
<td>2.1 6.1 11.3 18.1</td>
</tr>
<tr>
<td>Urology and Nephrology</td>
<td>1.00</td>
<td>188</td>
<td>2.0 5.6 10.4 18.7</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>0.99</td>
<td>157</td>
<td>1.8 4.8 9.6 19.0</td>
</tr>
<tr>
<td>Critical Care Medicine</td>
<td>0.99</td>
<td>61</td>
<td>1.3 3.4 10.0 17.2</td>
</tr>
</tbody>
</table>

**Policies and Practices of Agencies or Organizations that Fund or Conduct Systematic Reviews**
Respondents. Of the 22 Internet surveys sent by email request, 19 organizations responded yielding an overall response rate of 86%, with 17 groups having completed all mandatory questions. Responding organizations were from the U.S., Canada, U.K., and Australia. The majority of respondent organizations identified themselves as producers of systematic reviews (13/19; 68%) with the remainder presenting as both funder and producer combined (4/19; 21%), or exclusively as funders (2/19; 11%). Of those groups surveyed, all indicated they were not-for-profit, and were predominantly academic institutions (9/18; 50%) or national government agencies (5/18; 28%). Government research or infrastructure grants accounted for the majority of funding as reported by groups (16/18; 89%) followed by non-profit academic or non-governmental organization funding (8/18; 44%), internal funding (6/18; 33%) and industry or private sector funding (6/18; 33%).

Main findings. The majority of organizations indicated they produced systematic reviews for the collective goal of both knowledge and decision support (74%; 14/19), while 21% (4/19) reported producing reviews for decision-support. A large portion of respondents (15/19; 79%) view the importance of updating systematic reviews as high to very high. In spite of this however, most organizations do not have a policy in place for updating (13/19; 68%). Nevertheless, of these groups with no formal update processes, 54% (7/13) indicated establishing a policy was of importance. Of those organizations that reportedly update, 68% (13/19) indicate they do so irregularly. Approximately two thirds (13/19; 68%) of respondents reported that at least 20% of the reviews they commission or produce are out of date, and 32% respondents (6/19) reported that at least 50% of their reviews were out of date. When looking at issues of accountability, respondents specified that funder(s) of the original review (5/19; 26%), authors of the original review (5/19; 26%), and policymakers utilizing the evidence (3/19; 16%) were most responsible for ensuring systematic reviews are updated.

The use of formal methods to determine the need to update a systematic review was reported by 32% (6/19) of groups surveyed, 32% (6/19) reported the use of informal methods while an additional 37% (7/19) reportedly use no methods. When looking more in depth at updating strategies and practices, approximately half of organizations do not engage in regular literature searches to identify new evidence (10/19; 53%). However, of those groups that search periodically, searching frequencies were quite variable, with one group reporting monthly searching; two groups reporting every 12 months; one group indicating every two years; and one group stipulating that searching was dependent upon the stability of the evidence base and the relevance of the topic to their audience. The two most frequently reported strategies used (sometimes, often, or always) to monitor the emergence of new evidence were contacting experts in the field (14/18; 78%) and conducting general literature searches including electronic and hand searches (11/19; 58%). Additional surveillance strategies are listed below in  Table 8.

<table>
<thead>
<tr>
<th>Table 8. Monitoring Strategies</th>
<th>N of Respondents; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts in the field</td>
<td>14/18; 78%</td>
</tr>
<tr>
<td>General literature searches</td>
<td>11/19; 58%</td>
</tr>
<tr>
<td>Automatic database alerts or surveillance software</td>
<td>9/17; 53%</td>
</tr>
<tr>
<td>Systematic reviews surveillance</td>
<td>9/18; 50%</td>
</tr>
<tr>
<td>Guideline or health technology assessment surveillance</td>
<td>7/18; 39%</td>
</tr>
<tr>
<td>Trial registry surveillance</td>
<td>7/18; 39%</td>
</tr>
<tr>
<td>Statistical approaches</td>
<td>2/18; 11%</td>
</tr>
</tbody>
</table>
When examining updating influences, individuals or groups that reportedly impact most (sometimes, often or always) upon an organization’s decisionmaking process of whether to fund or conduct an update are as follows: external policymakers (16/19; 84%); the organization itself as the funder of the systematic review (15/19; 79%); experts in the field (13/19; 68%); and authors of the original review (13/19; 68%). Statisticians (1/19; 5%) and information specialists (3/19; 16%) were least likely to impact this decision. We also note that 26% of groups surveyed indicated that patients or consumer groups ‘sometimes’ influence this decisionmaking process.

When assessing specific issues that may factor into determining ‘when’ to update, a formal request from a policy or healthcare decisionmaker is the most frequently cited factor by the majority of respondents (16/19; 84%) followed by the totality of all new evidence under consideration (13/17; 76%). See Table 9 for additional impact factors.

Table 9. Factors that Impact on Determining “When” to Update

<table>
<thead>
<tr>
<th>Impact Factor</th>
<th>N of Respondents; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal request from a policy or healthcare decision maker</td>
<td>16/19; 84%</td>
</tr>
<tr>
<td>Totality (comprehensiveness) of all new evidence or data including harms &amp; benefits</td>
<td>13/17; 76%</td>
</tr>
<tr>
<td>Number of new studies identified</td>
<td>11/17; 65%</td>
</tr>
<tr>
<td>Reporting of serious or ‘new’ serious adverse events</td>
<td>11/17; 65%</td>
</tr>
<tr>
<td>Time credibility</td>
<td>10/18; 56%</td>
</tr>
<tr>
<td>Need for an internal organizational decision</td>
<td>8/16; 50%</td>
</tr>
<tr>
<td>New inclusion criteria (outcomes; interventions; populations; methodological advances/new analysis)</td>
<td>8/17; 47%</td>
</tr>
<tr>
<td>Number of participants in new studies</td>
<td>8/17; 47%</td>
</tr>
</tbody>
</table>

Additional updating influences include the notion that updating will have an effect on clinical practice (15/18; 83%), policy (13/18; 72%), organizational credibility of being current (13/18; 72%), current public controversy or interest (12/18; 67%), or cost utility of updating (12/18; 67%).

Data collected indicates that 60% (9/16; 56%) of respondents spend over 3 months of effort per review on activities related to updating systematic reviews, and 36% (6/16) reported expending over 6 months on updating. When looking closer at type of updating involvement, 72% of groups (13/18) report having ‘sometimes’ or ‘often’ been involved in doing full updates of all sections of a review. Two-thirds of respondents report ‘often’ or ‘sometimes’ having been involved in partial updates involving only certain sections of original reviews, while 61% of the groups (11/18; 59%) report having been involved in conducting an entirely new review upon updating. Only 1 of 18 respondents (6%) reported ever having discussed the need for a future update in the text of a systematic review. One third (5/17; 29%) of groups have withdrawn at least one systematic review from circulation after assessing the review as out of date. Approximately, 78% of organizations (14/18) reported they are ‘often’ or ‘sometimes’ able to draw on the same people involved in the original review. When asked if they had been involved in updating systematic reviews done by others, 61% (11/18) of respondents indicated they ‘seldom’ or ‘never’ done this, six groups ‘sometimes’ had, and only one group reported ‘often’ updating reviews done by others.
From the data gathered it would seem that most organizations are seldom or not utilizing current existing methods, such as cumulative meta-analytic approaches, when undertaking updating. The most frequently used approach is the time-based approach implying a pre-set updating frequency (7/18; 39%). (See Table 10.)
Identifying recent literature published after the date of the last search but before completion of the final systematic review is quite common among those surveyed with 94% (17/18) of organizations reporting this happens ‘sometimes’ or ‘always’. Organizations also report that this information is usually incorporated as an addendum in the review (11/18; 61%), or as a formal revision to the analysis (9/18; 50%).

**Updating Barriers.** Several elements of original systematic reviews were identified as moderate to serious barriers when updating as reported by respondents including the perceived need to redo data extraction (11/18; 61%); to change the original screening questions (9/18; 50%); to re-assess study quality (9/18; 50%) and to change the original search strategy (8/18; 44%). Further, respondents identified more broad-spectrum barriers (moderate to serious) to updating including limited funding and resources (17/18; 94%); limited academic credit for updating work (11/18; 61%); and limited publishing formats (9/18; 50%). With knowledge of the aforementioned barriers, it should be noted that 72% (13/18) of organizations reported knowing a systematic review was out of date but were not able to commence updating due to lack of resources (e.g. funding, personnel, time).

**Harmonization.** By harmonization we mean that different groups involved in the funding, conduct, or reporting of systematic reviews would come together and harmonize on issues of conduct, reporting and policy as it relates to updating systematic reviews. A large portion of respondents (11/19; 58%) indicated they ‘somewhat’ or ‘strongly’ support centralizing updating efforts across institutions or agencies that produce systematic reviews (i.e., harmonizing updating efforts). There were several perceived benefits (moderate to major) to participating in international harmonization efforts for updating with the foremost being the use of existing resources more efficiently (15/18; 83%). See Table 8 for a list of additional benefits.

**Table 11. Major/Moderate Benefits to Harmonization**

<table>
<thead>
<tr>
<th>N of respondents; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of existing resources more efficiently</td>
</tr>
<tr>
<td>Potential to minimize duplication of services</td>
</tr>
<tr>
<td>Access to new information, ideas, materials or other resources</td>
</tr>
<tr>
<td>Ability to address issues beyond a single organization’s domain</td>
</tr>
<tr>
<td>Share responsibility across organizations for complex/controversial issues</td>
</tr>
</tbody>
</table>

Respondents also indicated several barriers to harmonization, including the possible diversion of an organization’s funding resources (15/17; 88%) and insufficient human resources...
(14/17; 82%). As well, 76% of those surveyed (13/17) viewed perceived delays in working across organizations and possibly diverting the focus of research mandates within organizations (8/17; 47%) as moderate to serious barriers to collaboration. Obstacles aside, 84% (16/19) of the sample indicated they ‘somewhat’ to ‘strongly’ favored the development of a central registry of systematic reviews, which would be similar to efforts within the clinical trials community.
Chapter 4. Discussion

Among 100 systematic reviews, qualitative or quantitative signals for updating occurred for 57% (95% CI: 47%-67%) of the cohort. Median survival free of a signal for updating was 5.5 years (95% CI: 4.6-7.6). However, in 23 cases, a signal for updating occurred within 2 years, in 15 cases the signal occurred in less than 1 year, and, for 7 reviews, signals for updating had already occurred at the time of publication. Cardiovascular medicine, heterogeneity in the original review, and publication of a new trial larger than the previous largest trial were associated with shorter survival times, while inclusion of greater than 13 studies in the original review was associated with increased survival. However, no feature of the original review significantly predicted a signal for updating occurring within 2 years of publication. Using a search protocol combining PubMed Related Articles and a subject search limited with the optimized clinical query search filter for therapies identified almost all new signaling evidence with median screening burden of 71 new records per review.

Signals for updating occurred frequently and within relatively short timelines. While certain features were associated with shorter survival, prediction of the need to update a particular systematic review within specific time frames of interest (e.g., 2 years, as in the main analysis, or 1 or 3 years as checked in sensitivity analyses) does not appear feasible. It is worth emphasizing that this result is unlikely to change with further research. We tested all readily discernible features of systematic reviews with plausible relationships to the need for updating. We found several factors with statistically significant associations with shorter survival, including two with hazard ratios in the range of 2-3, magnitudes that would certainly be of interest epidemiologically. However, as recently highlighted in a discussion of prognostic tools, associations of this magnitude, despite being of epidemiological interest, generally do not give rise to useful prediction tools. The strength of association required for an epidemiological feature by itself to provide a screening test with useful sensitivity and specificity is orders of magnitude higher (i.e., the factor would need to confer a risk of approximately 200-300 fold). It is extremely unlikely that any features of the original systematic review—alone or in combination—would ever increase the risk of a signal for updating within 2 years to such an extent that these factors could usefully identify reviews in need of greater vigilance (i.e., with acceptable positive and negative predictive values). As such, surveillance of the literature for new evidence holds greater promise than relying on features of the original review to identify reviews likely to need updating within short time periods. A preliminary approach is proposed here (below), but may be refined through additional research.

We also evaluated the extent to which growth in the literature varied across broad areas of inquiry, as defined by clinical specialty (e.g., cardiovascular medicine, infectious diseases, obstetrics and gynecology, psychiatry), in order to determine if different clinical areas warranted greater attention with respect to updating. At this broad level, we found that, while the absolute number of new trials published each year does vary quite widely (from a low of 61 new RCTs per year in critical care medicine to a high of 1238 RCTs per year in peripheral vascular diseases), the doubling time for RCTs was surprisingly constant across fields. Linear growth results in ever-longer doubling times for an evidence base, which may bring stability to reviews where shifts in the direction of research are not a complicating factor, i.e., for reviews that could be updated through cumulative meta-analysis.
This combination of wide variation in absolute numbers of trials but minimal variation in rate of growth suggests that clinical fields probably vary in the number of reviews at risk for requiring an update, but that the risk per review does not differ dramatically as a function of clinical field. Importantly, this analysis of clinical fields involved only the rate of production of new evidence at a very broad level. In the cohort analysis, we did find that the specific field of cardiovascular medicine conferred shorter survival time. This finding may reflect features of the field other than the rate of growth, as the rate of 814 new RCTs per year for cardiovascular medicine fell approximately in the middle of the range of values seen across all 13 clinical areas. Number of publications is but one indicator of the amount of new information available. Its appeal it that it is easily counted. Number of new trials, number of patients (new participant ratio), and number of new events are other potentially relevant units of information for determining the need to update. Although the rate of new trial accrual in cardiovascular disease was ranked six among the clinical areas examined, the studies were large compared to other areas (Figure 2).

Alternatively, the similar rates of growth of the literature at the level of broad clinical areas may mask wide variations in growth rates for specific topics within a field. For instance, the rate of new trials per year in cardiology includes the rate for new RCTs in valvular heart disease, which is very small, as well as the rate for new RCTs related to acute coronary syndromes, which is quite high. In fact, in any of the broad clinical fields, there are likely specific topics that have much greater research activity than the average indicated by the broad field as a whole. Systematic reviews of such rapidly changing topics are likely to be challenging and resource-intensive to maintain.

**Practical Implications**

Many journals, including those with high impact factors, now routinely publish systematic reviews. However, publishing updates of systematic reviews presents challenges because journals, still largely print in format, only have a certain amount of space to provide for reporting systematic reviews. It is unclear whether they will devote any space for publishing updates. Although this situation might be less problematic for electronic journals, at least one of which is committed to publishing systematic review updates, journal ‘real estate’ available for updating is largely unknown. Print journals might consider publishing updates as “web-only” material. While this would avoid the problem of limited journal ‘real estate,’ it would still add to the peer review and editing workload for journals. In the case of AHRQ, the EPC program might want to consider developing its own peer reviewed, open access, indexed journal. Such a move might open an important dissemination venue for publishing systematic review updates.

Since its inauguration in 1997 the EPC program has already produced more than 150 evidence reports. And unlike other systematic reviews, which typically focus on a single question, EPC reports usually contain multiple systematic reviews within a single evidence report. Thus, EPC reports may require updating to an even greater extent than indicated by the present analysis, given the multiple topics addressed in each EPC report. With the development of the Medicare Modernization Act and the subsequent development of the clinical effectiveness reviews, and the renewal of the EPC program for another five years, there is likely to be a growing number of completed evidence reports. As with other systematic reviews, the utility of these evidence reports depends on their remaining up to date. Yet, most reviews are not kept up
to date.\textsuperscript{1} While it may be possible to commission each EPC with responsibility for keeping their completed evidence reports up to date, there are obvious drawbacks to such an approach. For example, principal authors of some reviews will not remain at the same centers, making updating more complicated. Competing demands on authors’ time as well as insufficient resources allocated to updating represent additional barriers.

Alternatively, the EPC program might commission one of its existing EPCs to centralize and harmonize the updating functions for the entire EPC program. Such an approach would likely be cost efficient and relieve pressures on individual EPCs to focus on identifying new evidence. Moreover, centralizing the updating process, an approach the Cochrane Collaboration is currently experimenting with, would facilitate efforts to study and improve the process of updating. This opportunity is important as the current evidence base to inform how and when to update systematic reviews is limited and new approaches and methodologies need to be developed. Through internal EPC knowledge translation activities such developments could be shared with other EPCs and the wider research and health policy community.

For Users of Systematic Reviews

Users of systematic reviews need to recognize that reviews can become out of date within relatively short time frames. Due to the peer review process, including sequential rejections at different journals, the time between the last electronic search and the time of publication can be considerable. Although our sample had an average ‘lag’ period (between the reported last search and eventual publication date) of a little more than 1 year, this average result reflects considerable variation. In fact, over 50% of the cohort had a publication lag of 1.4 years or greater and 25% had a lag of 2.1 years or greater.

To assess the degree to which a given systematic review remains up to date, readers should examine the most recent search date reported. If the search is over two years old, or if readers cannot ascertain this information from the report, then readers should seriously consider the possibility that the review is out of date. In such cases, readers might consider searching for a new systematic review or new primary studies that address the topic of interest. Another option consists of using secondary publication sources, such as the ACP Journal Club or Clinical Evidence, to identify recently published reviews or primary studies.

For Producers of Systematic Reviews

The finding that 7% of systematic reviews had signals for updating at the time of publication suggests that authors and publishers need to manage production and dissemination times more efficiently. Our data, drawn from a broad range of systematic reviews, including paper-based journal articles, Cochrane reviews, and health technology assessments, revealed that the median time from last search date to publication was 1.4 years, with a 25\textsuperscript{th} percentile of 0.9 years. However, among reviews that explicitly indicated that an update to the search was performed, the 25\textsuperscript{th} percentile was 28 weeks (0.5 years). Achievement of this benchmark across a greater proportion of systematic reviews would produce important increases in survival time.

It is unclear to what extent publishers of systematic reviews can expedite the peer review and publication process for systematic reviews any more than already attempted for submissions of
all types. However, authors can control two of the factors that contribute to publication lags. First, they can update searches prior to submission to capture any evidence that may have emerged during preparation of the review. Second, authors might consider submitting their work to the journals most likely to accept a given review in order to avoid delays due to multiple iterations of the peer review process. When the process of submission and rejection from other journals has resulted in the passage of 1 year or more since the date of the last search, authors should consider updating the search prior to resubmission.

 Proposed Surveillance Search Methodology

Searches constructed to conduct systematic reviews typically have low precision (often less than 5%) because they maximize recall in order to avoid missing any relevant evidence. Signal detection against the noise of very low precision makes standard systematic review searches inefficient for the purpose of identifying the need for updating. Good recall of the most influential new evidence would ideally be balanced with a low screening burden. Among the search strategies tested, the combination of two, Related Articles and the optimized Clinical Query, was sufficient to detect all relevant new studies indexed by the date of the updating signal in all but 4 cases. They achieved this recall with precision of approximately 35%. Assessing the performance of our methodology required that we screen candidates sequentially by indexing date in order to identify the earliest point at which a signal could be detected. Using this approach, we found that we needed to screen an average of 71 records per review to detect a signal for updating. In practice, however, there would be no requirement to screen purely in chronological order. For instance, one might first review citations retrieved from the top 5 general journals and the highest impact specialty journals related to the topic of interest. Triaging by journal source would likely reduce the ‘number needed to screen’ substantially.

In this retrospective assessment, reviewing newer systematic reviews was a useful adjunct to identifying new trials. This technique would have less utility for real time surveillance as lag of at least a year can be expected between trial publication and publication of a systematic review including that trial. Also, relying on newer systematic reviews means that the horse will already be out of the barn, as one would in effect be detecting the need for updating by identifying that an update had in fact been performed. Nonetheless, including both systematic reviews and RCTs adds little work, as the number of newer systematic reviews that require consideration will generally be small.
**Review Method**

Our approach consisted of applying explicit quantitative and qualitative criteria to new evidence. To ascertain quantitative criteria, we selected a limited number of major outcomes of interest and updated the meta-analytic estimates using fixed effects models. To ascertain qualitative criteria, we monitored for new evidence that changed the certainty of the previous findings, identified new harms or major caveats to the previous findings, among other clinically relevant changes in evidence. We did not compare this method to other possible approaches, but the approach has substantial face validity, and it proved feasible across the cohort of 100 systematic reviews.

As a starting point, we propose that major outcomes be identified for monitoring. The outcomes selected would be sufficiently central to the purpose of the review that signals for changed evidence would warrant undertaking an update. Surveillance searching in the form of the related article protocol and clinical queries would than be initiated, with a search frequency of twice yearly. Search results would than be screened and the robustness of findings from the original review be assessed in the face of the new evidence.

The method could be refined or tailored to different circumstances, including changing the way in which it is operationalized. Certainly the original team of authors for a given systematic review could carry out the same methodology. Alternatively, an agency that supports systematic reviews could carry out some of the surveillance and screening with a team of reviewers, but would need to involve one or more authors from the original review or others with expertise in the relevant content area.

**Review Frequency**

Agencies that fund systematic reviews should consider how often it is practical to conduct searches, particularly electronic searches, to identify potential new evidence for inclusion in a systematic review. We found that 4% of reviews had a signal for updating within 1 year of the last search date, and that 11% of reviews had signals for updating within 2 years of the last search date. Assuming that 4% is an acceptable risk and that 11% approaches an unacceptable risk, we suggest performance of the electronic searching process approximately once every six months. This frequency would allow sufficient time that, in the event new evidence is identified, reviewers would be able to carry out an update to their existing review along with publication within another 12 to 18 months, bringing the entire process to 2 years or less.

**Central or Distributed Surveillance and Updating**

For agencies or organizations that sponsor systematic reviews, the functions of surveillance, the decision to update, and actually performing the update could be central or distributed. A central approach would have an editorial group, a specialized team or administrative authority undertake these activities. A distributed approach would have these activities performed by the team that prepared the systematic review. Central surveillance is feasible, and a central approach may be better able to integrate factors other than new evidence, that may influence the decision
to update. Central surveillance permits a standard methodology that may facilitate priority setting when many reviews have signals for update.

Some updating functions, such as determination that a review is not in need of update, can be done centrally. When a major update is needed, it would most efficiently be performed by the authors of the original review.

Review authors could develop and submit an updating brief as part of the initial review – included and excluded files, explicit identification of the 3 newest and 3 largest included trials, and a summary of any information known to them at the time of submission that is pertinent to updating, such as trial registration numbers for ongoing trials.

**Format of Update**

Our results suggest that the risk factors we identified were not helpful in predicting signals for updating systematic reviews. Surveillance may prove to be more helpful although, as yet, it is unclear which of the several competing surveillance approaches might be most effective. Regardless of which method(s) will turn out to be most effective it is important to remember that these are only signals for updating. Deciding to ‘act upon’ an updating signal is likely to vary depending on several factors, such as priority setting by the funders or the interest in the topic. One approach to consider is to develop a decision tree model for each funder and/or producer. The details of such an approach are beyond the scope of the current project.

The flexibility and speed of amending electronic products, and the ability to retract obsolete versions makes that format attractive. Costs for both producers and users of electronic products are less than paper copies. Updating efforts should be restricted to the electronic version. Paper versions may be useful for initial dissemination to an important target audiences, however readers could be informed that only the electronic version is updated and so it should be considered the authoritative version.

**Surveillance Costs**

Life cycle costing of information products, including systematic review, would include updating and retirement costs as well as the initial costs of production and dissemination. In factoring updating into life cycle planning, not only the cost but also the potential to leverage the initial investment by extending the useful life of the systematic review should be considered. We are unable to isolate costs of various aspects of our assessment from costs associated with the methodological research, such as enhanced record keeping and extended search methods. The time requirements described below are estimates.

One experienced searcher with a Masters in Library and Information Science and technical competencies in record manipulation was able to provide abstracts of new evidence for assessment in one working day per systematic review. Time for subsequent surveillance searches on the same systematic reviews would be considerably less. Additional time is needed for provision of full text of articles appearing relevant, which in this cohort was a median of 7 articles (inter-quartile range; 4-24).

The reviewers who assessed the new evidence for updating signals were trained in medicine and clinical epidemiology. Depending on the complexity of the review, the initial orientation to the review and set up for calculation of quantitative signals required one half to one day of effort.
Time required to review candidate studies was dependent on the volume to be assessed. In this project, candidates spanned the period since the search of the original review and included the result sets of 5 search approaches that were under investigation. The median screening burden of the most effective combination of searches was 71 records per review (inter-quartile range 25 – 106) with a median of 7 obtained and assessed in full text format. Reviewing candidates and integrating relevant ones into the quantitative calculations required from one half to three days, but would be less in real-time surveillance. Validation with current expert opinion sources (which would not necessarily be available during real-time surveillance) and preparation of the case summary for review required several hours. Two to five systematic reviews could be considered in a two-hour team conference. Both preparation of the summary briefing and discussion at the team conference was more time consuming for reviews with signals than those without. With set up considered as a one-time cost, a review and summary might take 0.5 days for a review without a signal, and 1.5 days for one with a signal for updating.

Survey

The survey data collected indicate that most organizations that fund and/or conduct systematic reviews research consider updating as important. Unfortunately, most do not have updating policies in place. However, establishing updating procedures is viewed as noteworthy and something organizations concur should be considered. There is strong agreement that the proposed definition of ‘update’ is a valid explanation of the process, which over time will help to establish common nomenclature for this emerging methodological area.

The majority of survey respondents support the idea of harmonization of updating efforts across groups. Coordinating updating resources across organizations may facilitate performance of regular searches to identify new literature, something that currently occurs inconsistently at best, and may also foster development and evaluation of formal methods to determine the need for updating. Also, of note, most respondents (84%) favored the development of a central registry, analogous to efforts within the clinical trials community, to coordinate updating activities across agencies and review groups.

Even though a high response rate was achieved, the sample denominator was small, as this was a pilot survey. Therefore, interpretation of results should be made with caution. While these survey data will help to establish a baseline of current updating practices and policies of agencies that sponsor systematic reviews, we plan to undertake a similar survey with a larger sample of organizations engaged in funding or producing systematic reviews.

Limitations of the Review

We conducted the survival analysis using a retrospectively assembled cohort to determine how quickly systematic reviews require updating. The use of a retrospective cohort would be unlikely to bias the results of our analysis, but may have made the method appear more feasible than it would have appeared with a prospective cohort. The retrospective approach allowed us to use newer published systematic reviews, which aided in study identification for a substantial minority of cohort reviews. Relying exclusively on identification of new trials as they came out, as would occur with a prospective approach, would likely require greater effort. The
retrospective approach also broadened the availability of published expert opinion to validate our conclusions. Instead of just editorials accompanying the new trials, we could use information from newer textbooks, practice guidelines, and systematic reviews to guide our assessments of the impact of new evidence on the findings of the original review.

We chose systematic reviews indexed in *ACP Journal Club* as our sampling frame in order to construct a cohort of reviews of above average quality that were also directly relevant to clinical practice. It is possible that choosing *ACP Journal Club* introduced a bias in our cohort, such as preferential inclusion of reviews with positive results. A recent comparison of randomized controlled trials summarized in *ACP Journal Club* with the general population of trials indexed in MEDLINE found that *ACP Journal Club* preferentially abstracts randomized controlled trials that report positive results.24 *ACP Journal Club* might also preferentially focus on systematic reviews with positive results. While this could introduce some imbalance in the sample, such a bias would be unlikely to undermine our findings, as users of systematic reviews likely act on systematic reviews with positive findings more often than ones with null or negative findings. Nonetheless, understanding differences in updating between null reviews and those with clearly positive findings may represent a useful avenue for future methodological research.

We excluded from our analysis all qualitative reviews, reviews of non-therapeutic topics, individual patient data meta-analyses, and meta-regressions, based on our concerns that rates of change in evidence might differ across these different types of reviews. Restriction of our analysis to systematic reviews of randomized trials of conventional drugs, devices, or procedures that reported meta-analytic results for at least one dichotomous outcome may thus seem to have limited generalizability. However, as shown in Figure 2, excluding the records retrieved by our initial electronic search that were not systematic reviews, 139 (48%) of the first 287 systematic reviews were eligible for inclusion. (39 of the 139 eligible reviews were not included because they addressed topics that overlapped with earlier, eligible reviews, but they were nonetheless eligible for inclusion). Thus, while our cohort may appear highly selected, approximately half of the reviews indexed in *ACP Journal Club* were eligible for inclusion in our cohort, reflecting the fact that quantitative reviews of conventional drug therapies represent a substantial proportion of the systematic reviews directly relevant to clinical practice.

Our use of a structured approach for assessing differences between studies of the same topic without involving panels of experts represents the norm in methodological work of this type, including assessments of discrepancies between systematic reviews and large trials,25,26 variations in results between studies of different designs,14,27,28 differences in results presented in abstract form versus subsequent journal articles,12 and highly cited trials versus other trials addressing the same question.13 Nonetheless, assessments of the need to update previous systematic reviews would ideally include input by content experts who had evaluated the new evidence. As part of a follow-up project, we are conducting such an exercise with a subset of systematic reviews in this cohort to provide validation for the approach used. The idea of quantitative thresholds for indicating the need to update a previous analysis has face validity, but there is no basis for choosing a generic threshold. Therefore, we could have explored the impact on our results of different choices for these thresholds (e.g., using a threshold of 25% for changes in the magnitude of effect estimates, instead of 50%). Several alternatives were identified in a review of systematic review updating methods4 including optimal information size29 and new participant ratio.5 Signals based on these quantities could be developed for real world updating in place of or in addition to these current criteria. In the case of an individual review, such
approaches, especially optimal information size, would be attractive as they would be less arbitrary. However, operationalizing these methods over a cohort of some 100 reviews would require making a number of imputations that would themselves be somewhat arbitrary (e.g., generic values for event rates in control groups and expected effect sizes).

The idea of qualitative changes in evidence that warrant updating a previous review also has face validity, but, again, the specific signals we chose were somewhat arbitrary. However, the concepts captured by our qualitative signals (substantial changes in certainty, new harm, emergence of superior alternative treatments, important caveats about the patient populations who benefit from treatment, and other such issues) emerged from input from our technical advisory panel, as well a published framework for evaluating the need to update clinical practice guidelines. Others might modify the specific criteria we chose, but we believe the qualitative criteria we used speak for themselves as representing clinically relevant changes in evidence.

Finally, this report presents several novel lines of enquiry concerning updating systematic reviews. We have stated what we believe to be logical implications of our findings, but it will be important to have others groups replicate of research and see whether such replication results in findings similar to ours. This is a rich area for research and as well as independently replicating, we strongly encourage others to extend and refine these lines of research.

Conclusions

In a cohort of high quality systematic reviews directly relevant to clinical practice, signals for updating occurred frequently and within relatively short timelines. A number of features significantly affected survival, but none significantly predicted the need for updating within 2 years.

Methods for identifying reviews in need of updating based on surveillance for new evidence hold more promise. Several of the methods tested were feasible, yielding good recall of relevant new evidence with modest screening burdens.

The majority of organizations engaged in the funding or production of systematic reviews view the importance of updating systematic reviews as high to very high. Despite this recognition, most organizations report having no formal policy in place for updating previous systematic reviews. Slightly less than half of organizations performed periodic literature searches to identify new evidence, but searching frequencies varied widely, from monthly to every two years.

If systematic reviews are to achieve their stated goal of providing the best evidence to inform clinical decision making and healthcare policy, issues related to identifying reviews in need of updating will require much greater attention. In the meantime, publishers of systematic reviews should consider a policy of requiring authors to update searches performed over 12 months prior to submission. And, users of systematic reviews need to recognize that important new evidence can appear within short timelines. When considering the results of a particular systematic review, users should search for more recent reviews or trials to see if any exist and determine if the results are consistent with the previous review.
References and Included Studies

Cohort Meta-Analyses Sorted by Signals

Primary signal for updating

Potentially invalidating qualitative signal (with or without quantitative signal)


Major signal for updating


**Quantitative signal involving primary outcome but no qualitative signal**


**Any signal**

Quantitative signal involving non-primary outcome or any of three pre-defined criteria for changes in the amount or precision of information: increase in total number of trials by ≥ 50%, increase in total number of participants by ≥ 50%, publication of a new trial with sample size ≥ 3 times size of previous largest trial, relative of change ≥ 50% for the width of the 95% confidence interval.


77. Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. Cochrane Database Syst Rev. 2004(2):CD000363.


New trials but no signal


89. Heidenreich PA, McDonald KM, Hastie T, et al. Meta-analysis of trials comparing beta-


**No new trials**


References


23. Many reviews are systematic but some have better systems than others [Editorial]. PLoS Med 2007; 4(3)


24. Crowley P. Interventions to prevent, or improve outcome from, labour at or beyond term. Cochrane Database of Systematic Reviews 1997;(2):Focused on ineligible therapy.


48. Ebrahim S, Davey Smith G. Multiple risk factor interventions for primary prevention of coronary heart disease. Cochrane Database of Systematic Reviews 1999;Not clinical benefit or harm of a specific (class of) drug, device or procedure.


Fraser W D, Krauss I, Brisson-Carrol G et al. Amniotomy to shorten spontaneous labour. Cochrane Database of Systematic Reviews 1995;(2):Date of original could not be established.


Golden R N, Gaynes B N, Ekstrom R D. The efficacy of light therapy in the treatment of mood


Glennie J, for the, Canadian Coordinating et al. The efficacy of tacrine and the measurement of outcomes in Alzheimer's Disease. CCOHTA Technology Overview: Pharmaceuticals 1997;5No MA for outcome in abstract.


100. Herxheimer A, Petrie K J. Melatonin for preventing and treating jet lag. Cochrane Database of Systematic Reviews 2001;(1): Date of original could not be established.


123. Jones A, Fay J K, Burr M. Inhaled corticosteroid effects on bone metabolism in asthma and mild chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2002;(1):Summary statistic was continuous, not WMD.


125. Julian D G, Camm A J, Frangin G et al. Randomised trial of effect of amiodarone on...


150. Levine C B, Fahrbach K R, Siderowf A D. Diagnosis and treatment of Parkinson's disease: a systematic review of the literature. Evid Rep Technol Assess (Summ) 2003;571-4. Summary statistic was continuous, not WMD.


152. Low D E, Desrosiers M, McSherry J. A practical guide for the diagnosis and treatment of acute sinusitis. CMAJ 1997;156(Suppl. 6):1S-14S. Focused on ineligible therapy.


187. Miller R G, Mitchell J D, Moore D H. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database of Systematic Reviews 1999;Date of original could not be established.


O'Connor A M, Rostom A, Fiset V. Decision aids for patients facing health treatment or screening decisions: systematic review. BMJ 1999;319:731-4. Not clinical benefit or harm of a specific (class of) drug, device or procedure.


## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ</td>
<td>Agency for Healthcare Research and Quality</td>
</tr>
<tr>
<td>CCJ RCT</td>
<td>Abbreviation use for one of the search strategies tested. CCJ is the Core Clinical Journals, a MEDLINE journal subset previously known Abridged Index Medicus.</td>
</tr>
<tr>
<td>CCT</td>
<td>Controlled clinical trial</td>
</tr>
<tr>
<td>CQ</td>
<td>Abbreviation use for one of the search strategies tested - Clinical Queries</td>
</tr>
<tr>
<td>CR</td>
<td>Abbreviation use for one of the search strategies tested - Citing References</td>
</tr>
<tr>
<td>EPC</td>
<td>Evidence-Based Practice Center</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>HTA</td>
<td>Health Technology Assessment</td>
</tr>
<tr>
<td>ISI Science Citation</td>
<td>ISI is Institute for Scientific Information</td>
</tr>
<tr>
<td>IRQ</td>
<td>Inter-quartile range</td>
</tr>
<tr>
<td>MeSH</td>
<td>Medical Subject Headings – the term describes the MEDLINE thesaurus</td>
</tr>
<tr>
<td>PMID</td>
<td>PubMed unique identifier</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>RA RCT</td>
<td>Abbreviation use for one of the search strategies tested – Related Article RCT</td>
</tr>
</tbody>
</table>
APPENDIXES
Appendix A. Definitions and Criteria for Signals for Updating

A. Qualitative signals for need to update

We defined signals for two categories of qualitative signals for potential changes in evidence (i.e., for the need to update the original meta-analysis) in terms of their level of importance.

Potentially invalidating change in evidence: one would no longer want clinicians to act upon the results of the original review; an agency or organization that supported the production of the original review would want to retract the review until it could be updated. Examples of such changes include: high quality new evidence that suggests conclusions opposite to those in the original review; high quality new evidence suggests a degree of harm that would completely undermine use of the therapy; or, a head-to-head trial data show that the treatment evaluated in the original review is substantially inferior to another treatment. The specific operational details for each of the three criteria for potentially invalidating changes in evidence are provided below. Importantly, the designation ‘potentially invalidating’ refers to the recommendations for clinical practice implied by the original meta-analysis, not the methods or conduct of the meta-analysis itself.

Major change in evidence: the conclusions of the original review have not been overturned or superseded, but new evidence clearly has the potential to affect clinical decision-making. Examples of such changes include: new evidence that suggests the therapy does not work in certain patient populations; new evidence that affects how the therapy must be delivered in order to confer the benefit suggested in the original review (e.g., duration of treatment or in conjunction with other co-treatment); evidence about harm that would not completely undermine use of the therapy, but would clearly affect the decision to recommend therapy for at least some patient populations; changes in conclusion that fall short of ‘opposite’ but to those in the original review; high quality new

Qualitative signals were detected using explicit criteria for comparing the language used to characterize findings in the original meta-analysis with descriptions of findings in new meta-analyses that addressed the same topic, new ‘pivotal trials’, new clinical practice guidelines, or new editions of major textbooks (e.g., UpToDate). Pivotal trials were defined as trials that had a sample size at least three times the previous largest trial or were published in one of the 5 top general medical journals (New England Journal of Medicine, Lancet, Journal of the American Medical Association, Annals of Internal Medicine, and the British Medical Journal) based on a ranking by journal impact factor. Specific types of qualitative signals are defined below.

Criteria for Signals of Potentially Invalidating Changes in Evidence

A1. Opposing findings: Pivotal trial, meta-analysis including at least one new trial, practice guideline (from major specialty organization or published in peer-reviewed journal), or recent textbook (e.g., UpToDate) characterizes the treatment in opposite terms to those in the cohort
review: e.g., definitely effective → ineffective or vice versa (i.e., ineffective → effective). We operationalized ‘opposite’ as described at the end of this section.

We included guidelines and textbooks as sources of qualitative criteria because our definition of pivotal trial sets a very high bar. For example, we have not included any high impact specialty journals. The only way for a trial not published in one of the top 5 general medical journals to count as a pivotal trial would be for it to have a sample size at least three times that of the previous largest trial. To minimize our overlooking important new evidence while still permitting the efficiency of narrow searches for pivotal trials, we included guidelines and textbooks as sources of qualitative signals for changes in evidence. If new evidence has appeared that is judged of sufficient quality to inform recommendations in practice guidelines or textbooks, then it seems reasonable to call attention to these recommendations as signals for the need to update the original systematic review.

**A2. Substantial harm:** Pivotal trial, meta-analysis including at least one new trial, practice guideline, recent textbook calls into question the use of the treatment on the basis of harm (i.e., the treatment would no longer be recommended because risks outweigh benefits). A new result for harm that does not undermine use altogether, but has clear potential to affect clinical decision making would count as a ‘major change’ (criterion A6, ‘Important caveat’, as defined below).

**A3. Superior new treatment:** Pivotal trial, systematic review including at least one new trial, practice guideline, or recent textbook characterized another treatment as significantly superior to the one evaluated in the original meta-analysis (based on efficacy or harm)—to the point that it would be preferred in most settings.

**Criteria for Signals of Major Changes in Evidence**

**A4. Important changes in effectiveness short of ‘opposing findings’:** Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook does not contradict the previous review, but characterizes benefit in substantially different terms (e.g., therapy previously characterized as “promising”, “likely beneficial” or similar description and now characterized as definitely beneficial.) This criterion is defined below in greater detail in the explanation of ‘Operational definition of changes in conclusions.’ Importantly, no attempt was made to distinguish between varying descriptions of “possibly effective.” Characterizations such as “may be effective,” “promising,” “trends towards effectiveness,” and other similar phrases or concepts were all categorized as “possibly effective.” Thus, this criterion captured substantive differences in the characterization of treatment effects, not merely semantic differences.

**A5. Clinically important expansion of treatment:** Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook has expanded of the role of the treatment (e.g., the treatment has now been shown to be of benefit in children or the elderly; or benefit now shown to apply to primary prevention of disease, not just secondary prevention).

**A6. Clinically important caveat:** Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook adds an important caveat, about the patient populations who benefit, way in which treatment has to be delivered in order to derive benefit, sustainability of benefit (e.g., benefits on short term outcomes, but not long-term ones), or increases in harm that are not sufficient to undermine use altogether, but would clearly affect the decision to recommend treatment for at least some patient populations.
A7. Opposing findings from discordant meta-analysis or non-pivotal trial: The treatment has been characterized in sufficiently different terms to the cohort review that disagreement would have met criteria for ‘opposing findings’ (criterion A1) except the source was not a pivotal trial, new-meta-analysis, or more recent practice guideline, or recent textbook—rather, it was a discordant meta-analysis or trial indexed in ACP Journal Club. (‘Discordant meta-analysis’ was defined as one that reached different conclusions than the original meta-analysis, despite effectively covering the same search period.)

We included this criterion because our definition of pivotal trial sets a very high bar, including only the top 5 general medical journals and trials with sample sizes at least three times the size of the previous largest trial. This criterion allows other sources of evidence to count as qualitative singles, without allowing any new trial with different results than in the previous systematic review to count as a signal for updating.

Operational definition of changes in conclusions

Labels such as ‘effective’ and ‘ineffective’ do not capture the distinction between trends towards effectiveness or uncertainty in the face of conflicting results or major limitations of the existing evidence. On the other hand, attempting to capture such nuances runs the risk of regarding semantic or stylistic differences between different authors. To balance these concerns, we consider conclusions about effectiveness in terms of a 5-point scale as shown below.

<table>
<thead>
<tr>
<th>Definitely Effective</th>
<th>Probably/possibly Effective</th>
<th>Uncertain Effectiveness</th>
<th>Probably/possibly Ineffective</th>
<th>Definitely Ineffective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For systematic reviews that focused on adverse effects of treatment, we replaced effective/ineffective with ‘harmful/not harmful.’

In the interest of having qualitative signals of changes in evidence with high specificity, we did not attempt to make distinctions between statements of ‘probable’ or ‘possible’ benefit (or lack of benefit). We assigned descriptions such as ‘promising,’ ‘possibly’, ‘probably’, ‘maybe’, ‘likely’ into the same category. While still subjective, our labels are thus quite conservative—we are distinguishing firm or confident results, from trends, and equipoise or complete uncertainty (the middle position).

We defined ‘opposite’ conclusions (criterion A1 for potentially invalidating changes in evidence) as a movement of at least two positions on the above scale and ‘important changes in effectiveness short of opposing findings’ (criterion A4 for major changes in evidence) as a movement of one position on this scale. A movement of two positions generally includes movements from benefit to lack of benefit (or vice versa), but also includes movements from uncertain to definite conclusions about effectiveness. In this context, it is important to emphasize that we were careful not to equate summary conclusions (e.g., in article abstracts) of the type ‘the evidence does not permit definite conclusions’ with ‘complete uncertainty.’ In many such
cases, the results reported in the trial or meta-analysis indicated a trend, but the authors regarded the trend as inconclusive, on statistical or methodological grounds. Such cases were judged as ‘possible’ benefit (or lack of benefit, depending on the results). We reserved ‘completely uncertain’ for cases in which the authors clearly regarded the evidence as not indicating towards either benefit or lack of benefit. Thus, we regarded that a change from a definite or confident conclusion to compete uncertainty (or vice versa) would represent a potentially invalidating change in evidence. For example this would include a change from there being no basis on which to recommend a treatment to its being definitely recommended (or vice versa).

**B. Quantitative signals of changes in evidence**

We performed updated meta-analyses that combined the results from new trials with the meta-analytic result reported in the original review. To count as a quantitative signal, an outcome explicitly identified as a primary outcome in the original meta-analysis or any mortality outcome had to meet one of the criteria below. To count as a primary outcome, we required use of the word ‘primary’ or ‘main.’ Even in cases where those words were used, we discounted such outcomes if authors stated that they had more than 3 such outcomes (on the grounds that more than 3 undermines the concept of ‘primary’).

**B1. Change in statistical significance:** at least one of the 95% confidence limits lies on a different side of the line of no effect (i.e. odds ratio or relative risk=1, risk difference=0). This criterion captures whether a result that was statistically significant in the original systematic review is now not statistically significant or vice versa—a previously non-significant result has become statistically significant.

To avoid counting trivial or ‘borderline’ changes in statistical significance as quantitative signals for updating, we required that at least one of the two results (i.e., the original and updated meta-analyses) have a p-value outside the range of 0.04 to 0.06. In other words, we excluded cases in which the original systematic review reported a borderline result and the updated result is also borderline but happens to lie in the other side of the line of no effect. For instance, a change from \( p = 0.041 \) to \( p = 0.059 \) would not count as a quantitative signal to update, nor would the converse change (from \( p = 0.059 \) to \( p = 0.041 \)).

**B2. Change in effect size of at least 50%:** the new result indicates a relative change in effect size of at least 50%. For example, if \( \text{RRR}_{\text{new}} / \text{RRR}_{\text{old}} \leq 0.5 \) or \( \text{RRR}_{\text{new}} / \text{RRR}_{\text{old}} \geq 1.5 \), where RRR is the relative risk reduction. Thus, if the original review has found RR=0.70 for mortality, this implies RRR of 0.3. If the updated meta-analytic result for mortality were 0.90, then the updated RRR would be 0.10, which is less than 50% of the previous RRR. In other words the reduction in the risk of death has moved from 30% to 10%. The same criterion applied for odds ratios (e.g., if previous OR=0.70 and updated result were OR=0.90, then the new reduction in odds of death (0.10) would be less 50% of the magnitude of the previous reduction in odds (0.30). For risk differences and weighted mean differences, we applied the criterion directly to the previous and updated results (e.g., \( \text{RD}_{\text{new}} / \text{RD}_{\text{old}} \leq 0.5 \) or \( \text{RD}_{\text{new}} / \text{RD}_{\text{old}} \geq 1.5 \).
Appendix B: Examples of Systematic Reviews with Qualitative Signals for Potentially Invalidating Changes in Evidence

I. All 8 Reviews with Signals for “Potentially Invalidating Changes in Evidence” (criteria for signals A1-A3)


Question(s) addressed
Does human albumin or plasma protein fraction reduce mortality in patients who are critically ill (hypovolemia, burns, or hypoalbuminemia)?

Findings of Original Review
The original review found that “For each patient category the risk of death in the albumin treated group was higher than in the comparison group… an increase in the risk of death of 6% (3% to 9%). These data suggest that for every 17 critically ill patients treated with albumin there is one additional death.”

New Findings
A pivotal trial found no difference in the risk of death between patients who received albumin and those who did not (relative risk of death, 0.99; 95 percent confidence interval, 0.91 to 1.09; p =0.87). It concluded: “In patients in the ICU, use of either 4 percent albumin or normal saline for fluid resuscitation results in similar outcomes at 28 days.”

Qualitative signal
Opposing findings: The original systematic review reported an increase in mortality; a pivotal trial showed no difference in risk of death

Quantitative signal
Change in statistical significance: Relative risk of death became non-significant
RR = 1.68 (1.26, 2.23) → 1.04 (0.95, 1.13)

Change in effect magnitude of 50% or more: Relative risk increase for death of 0.68 → increase of only 0.04

Other signals
Increase in number of patients of at least 50%: N=1419 → N=8352
Trial with sample size at least 3 times the size of previous largest trial:
Previous largest trial had 219 patients; new trial had 6933 patients
Change in width of 95% confidence interval of at least 50%: as shown above

Source of Pivotal Trial

Time to signal  Qualitative signal: 3.0 years  Quantitative signal: same


Question(s) addressed Does intravenous immunoglobulin (IVIG) reduce mortality, bacteriological failure rates, and duration of stay in hospital in patients with bacterial sepsis septic shock?

Findings of Original Review Comparing polyclonal IVIG versus control, the original review reported a relative risk of death 0.60 (95% CI: 0.47 to 0.76) among a total of 413 patients. The authors concluded that polyclonal intravenous immunoglobulin “significantly reduces mortality and can be used as an adjuvant treatment for sepsis and septic shock.”

New Findings A subsequent meta-analysis (Pildal 2004) included 763 patients and found that "[h]igh-quality trials …showed a relative risk of 1.02 (95% CI, 0.84-1.24), whereas other trials (involving a total of 948 patients, 292 of whom died) showed a relative risk of 0.61 (95% CI, 0.50-0.73). Because high-quality trials failed to demonstrate a reduction in mortality, polyclonal immunoglobulin should not be used for treatment of sepsis except in randomized clinical trials."

The textbook Up-To-Date quotes this subsequent meta-analysis and states intravenous immunoglobulin “is rarely used to treat patients with septic shock in the United States, and this approach is not recommended pending the demonstration of benefit in large, well designed trials."

Qualitative signal  Opposing findings: The original systematic review reported a definite reduction in mortality; a subsequent meta-analysis showed no benefit

Quantitative signal  Change in statistical significance: among higher quality trials only
Change in effect magnitude of 50% or more: among higher quality trials only

Other signals: Increase in number of patients of at least 50%: N=1992 → N=3082 (this increase was for meta-analysis of monoclonal anti-endotoxins; for polyclonal IVIG, increase was not 47%)

Question(s) addressed?

What effects does calcium supplementation during pregnancy have on blood pressure, preeclampsia, and adverse maternal and fetal outcomes

Findings of Original Review

The original review showed a substantial, statistically significant reduction in the occurrence of preeclampsia among women who received calcium supplementation compared with placebo was (OR of 0.38; 95% CI, 0.22 to 0.65), as well as significant improvements in blood pressure. It concluded: "Calcium supplementation during pregnancy leads to an important reduction in systolic and diastolic blood pressure and preeclampsia."

New Findings

A pivotal trial published the following year reported: “Calcium supplementation did not significantly reduce the incidence or severity of preeclampsia or delay its onset… There were no significant differences between the two groups in the prevalence of pregnancy-associated hypertension without preeclampsia (15.3 percent vs. 17.3 percent) or of all hypertensive disorders (22.2 percent vs. 24.6 percent). The mean systolic and diastolic blood pressures during pregnancy were similar in both groups.” It concluded that “Calcium supplementation during pregnancy did not prevent preeclampsia, pregnancy-associated hypertension, or adverse perinatal outcomes in healthy nulliparous women.”

Qualitative signal

Opposing findings: The original systematic review reported a definite reduction in pre-eclampsia and development of hypertension; a pivotal trial showed no impact on either outcome.

Quantitative signal

Change in effect magnitude of 50% or more: reduction in odds of pre-eclampsia of 0.62 → reduction of only 0.21; and reduction in odds of developing hypertension of 0.70 → reduction of only 0.25

Other signals:

Increase in number of patients of at least 50%: N=2280 → N=7059

Trial with sample size at least 3 times the size of previous largest trial:

Previous largest trial included 1167 patients; new trial included 4779 patients

Source(s) of new evidence

Pivotal trial:

**Additional trials:**


**Time to signal**

Qualitative signal: 1.2 years
Quantitative signal: same


**Question(s) addressed**

In patients with carotid stenosis, what are the risks and benefits of endovascular treatment compared with carotid endarterectomy?

**Findings of Original Review**

The original review found no significant difference in the odds of treatment related death or any stroke (odds ratio [OR], endovascular surgery, 1.33; 95% confidence interval [CI], 0.86 to 2.04), death or disabling stroke (OR, 1.22; CI, 0.61 to 2.41), or death, any stroke, or myocardial infarction (OR, 1.04; CI, 0.69 to 1.57). At 1 year after randomization, there was no significant difference between the 2 treatments in the rate of any stroke or death (OR, 1.01; CI, 0.71 to 1.44).

It concluded: "No significant difference in the major risks of treatment was found but the wide confidence intervals indicate that it is not possible to exclude a difference in favor of one treatment. Minor complication rates favor endovascular treatment."

**New Findings**

One pivotal trial (Mas 2006) was stopped early because of significantly inferior outcomes for endovascular treatment. “The 30-day incidence of any stroke or death was 3.9% after endarterectomy (95% CI: 2.0 to 7.2) and 9.6% after stenting (95% CI: 6.4 to 14.0); the relative risk of any stroke or death after stenting as compared with endarterectomy was 2.5 (95% CI: 1.2 to 5.1).” Rates of death and stroke at 6 months were also lower with endarterectomy than with stenting.

Another pivotal trial (Ringleb 2006) found that “The rate of death or ipsilateral ischemic stroke from randomization to 30 days after the procedure was 6.84% with carotid-artery stenting and 6.34% with carotid endarterectomy (absolute difference 0.51%, 90% CI –1.89% to 2.91%). Based on a pre-defined non-inferiority margin of 2.5%, the authors concluded that endovascular treatment “failed to prove non-inferiority of carotid-artery stenting compared with carotid endarterectomy… The results of this trial do
not justify the widespread use in the short-term of carotid-artery stenting for treatment of carotid-artery stenoses.”

**Qualitative signal**

**Opposing findings:** The original review reported no major differences between the two treatments. The review emphasized the uncertainty of the comparison, but did not specifically indicate any possibility that endovascular treatment was inferior to endarterectomy. Two pivotal trials indicate inferiority of endovascular treatment (in one case, of sufficient magnitude to result in termination of the trial).

Editorials for both pivotal trials discuss possible explanation for these findings that leave open the possibility of non-inferiority. But the point remains that the publication of these two high profile trials with results substantially different from those of previous trials constitutes an important signal for the need for updating the original systematic review.

**Quantitative signal**

**Change in statistical significance:** Relative risk of stroke or death within 30 days became statistically significant, with both limits of 95% confidence interval now lying on side of increased risk with endovascular treatment

\[ \text{RR} = 1.33 \ (0.86, 2.04) \rightarrow 1.35 \ (1.02, 1.80) \]

**Other signals:**
- Increase in number of trials of at least 50%: 6 trials → 9 trials
- Increase in number of patients of at least 50%: N=1269 → 3376

**Source(s) of new evidence**

**Pivotal trials:**


**Additional trial:**


**Time to signal**

Qualitative signal: 1.5 years
Quantitative signal: same

**Question(s) addressed**

Is hormone replacement therapy (HRT) associated with cardiovascular events or cancer in postmenopausal women?

**Findings of Original Review**

The original review concluded that was no clear evidence of an association between cardiovascular outcomes and HRT, but noted that “Data on cardiovascular events and cancer were usually given incidentally, either as a reason for dropping out of a study or in a list of adverse effects.” We therefore characterized the original systematic review as having concluded that effectiveness was ‘uncertain’.

**New Findings**

A pivotal trial (Hulley 1998) found no difference between HRT and placebo in terms of the primary or secondary cardiovascular endpoints. (RR=0.99; 95% CI: 0.80 to 1.22). The trial also showed an increase in thromboembolic events. It concluded: “Based on the finding of no overall cardiovascular benefit and a pattern of early increase in risk of CHD events, we do not recommend starting this treatment for the purpose of secondary prevention of CHD.”

A second, larger pivotal trial (Rossouw 2002) was stopped early “because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits.” Based on a mean follow-up of 5.2 years, “[a]bsolute excess risks per 10 000 person-years attributable to estrogen plus progestin [HRT] were 7 more CHD events, 8 more strokes, 8 more PEs, and 8 more invasive breast cancers, while absolute risk reductions per 10 000 person-years were 6 fewer colorectal cancers and 5 fewer hip fractures. The absolute excess risk of events included in the global index was 19 per 10 000 person-years. “Overall health risks exceeded benefits from use of combined estrogen plus progestin for an average 5.2-year follow-up among healthy postmenopausal US women. All-cause mortality was not affected during the trial.

The risk-benefit profile found in this trial is not consistent with the requirements for a viable intervention for primary prevention of chronic diseases, and the results indicate that this regimen should not be initiated or continued for primary prevention of CHD.”

**Qualitative signal**

**Opposing findings**: The original systematic review found no clear relationship between HRT and cardiovascular outcomes. Two pivotal trials clearly demonstrated a lack of benefit and evidence of some harm.

**Quantitative signal**

**Change in statistical significance**: odds of increased cardiovascular and thromboembolic events became statistically significant.
Odds ratio of 1.64 (0.65, 4.18) \(\rightarrow\) 1.70 (1.18, 2.43)

Increase in number of patients of at least 50%: N=4124 \(\rightarrow\) 25140

Trial with sample size at least 3 times the size of previous largest trial:
previous largest trial had N=1265; new trial had N=16608
Change in width of 95% confidence interval of at least 50%: as shown above

Other signals:

**Source(s) of new evidence**


**Time to signal**

Qualitative signal: 1.1 years
Quantitative signal: same

---


**Question(s) addressed**

How efficacious and safe is interferon alfa with or without ribavirin in the treatment of chronic hepatitis C?

**Findings of Original Review**

The original review found that, compared with interferon alone, “combination therapy reduced the risk of not having a sustained virological for 6 months by 26% in naive patients (relative risk 0.74, 95% confidence interval 0.70 to 0.78), 33% in relapers (0.67, 0.57 to 0.78), and 11% in non-responders (0.89, 0.83 to 0.96). Morbidity and mortality showed a non-significant trend in favour of combination therapy (Peto odds ratio 0.45, 0.19 to 1.06). Combination therapy significantly reduced the risk of not having improvement in results of histology by 17% in naive patients (0.83, 0.74 to 0.93) and by 27% in relapers and non-responders (0.73, 0.66 to 0.82). The authors concluded that “treatment with interferon alfa plus
New Findings

Two pivotal trials compared the combination evaluated in the original systematic review with an alternative treatment, peginterferon alfa combined with ribavirin.

The first trial included three treatment arms, standard interferon alfa-2b plus ribavirin (as evaluated in the original review), pegylated interferon alfa-2b (1.5 µg/kg per week for four weeks followed by 0.5 µg/kg per week) plus ribavirin, and pegylated interferon alfa-2b (1.5 µg/kg per week) plus ribavirin. The primary endpoint of sustained virologic response “was significantly higher (p =0.01 for both comparisons) in the higher-dose peginterferon group (274/511 [54%]) than in the lower-dose peginterferon (244/514 [47%]) or interferon (235/505 [47%]) groups.”

They concluded “In patients with chronic hepatitis C, the most effective therapy is the combination of peginterferon alfa-2b 1.5 microg/kg per week plus ribavirin,” though they noted that “The benefit is mostly achieved in patients with HCV genotype 1 infections.”

The second pivotal trial (Fried 2002) found that “a significantly higher proportion of patients who received peginterferon alfa-2a plus ribavirin had a sustained virologic response (defined as the absence of detectable HCV RNA 24 weeks after cessation of therapy) than of patients who received interferon alfa-2b plus ribavirin (56 percent vs. 44 percent, p <0.001) or peginterferon alfa-2a alone (56 percent vs. 29 percent, p <0.001).” They concluded: “In patients with chronic hepatitis C, once-weekly peginterferon alfa-2a plus ribavirin was tolerated as well as interferon alfa-2b plus ribavirin and produced significant improvements in the rate of sustained virologic response, as compared with interferon alfa-2b plus ribavirin or peginterferon alfa-2a alone.”

The textbook Up-To-Date cites these two trials (and a subsequent trial that evaluated the optimal dose of ribavirin) in making the statement that “combination therapy with pegylated interferon plus ribavirin is generally associated with a higher sustained virologic response rate compared to combination therapy with standard interferon plus ribavirin or pegylated interferon monotherapy. As a result, this is usually the preferred approach in patients with hepatitis C who have not previously received treatment.” The chapter in Up-To-Date noted the influence of genotype on response, which was seen in both trials. Because the benefit in the first trial was largely confined to patients with a particular genotype, we did not take that trial by itself as the basis for the signal of a superior alternate treatment. We regarded the signal as triggered by the second trial (Fried 2002).

Qualitative Superior new treatment: Head to head comparisons in two pivotal trials
**signal** showed that an alternative treatment is superior to the therapy evaluated in the original systematic review.

**Quantitative signal** Not applicable – comparisons in new trials differ from those in the original systematic review

**Other signals:** None

**Source(s) of Pivotal trials:**


**Time to signal**

- Qualitative signal: 0.9 years
- Quantitative signal: Not applicable

---


**Question(s) addressed** Does the use of corticosteroids in patients with sepsis or septic shock lower the risk of death?

**Findings of Original Review** The original review found that “Corticosteroids did not change 28 day mortality (15 trials, n = 2022; relative risk 0.92, 95% confidence interval 0.75 to 1.14) or hospital mortality (13 trials, n = 1418; 0.89, 0.71 to 1.11).” The authors concluded that “No overall beneficial effect of corticosteroids in patients with septic shock was observed…”

**New Findings** A randomized, double-blind, multi-center trial evaluated the impact of a 7-day course of low-dose hydrocortisone versus placebo in patients who showed signs of relative adrenal insufficiency. It found a significantly lower risk of death in the corticosteroid group (hazard ratio, 0.67; 95% confidence interval, 0.47-0.95; p =0.02). It concluded that “a 7-day treatment with low doses of hydrocortisone and fludrocortisone significantly reduced the risk of death in patients with septic shock and relative adrenal insufficiency without increasing adverse events.”

A subsequent meta-analysis showed that, among five trials (n = 465) involving long courses (> or = 5 days) with low dose (< or = 300 mg hydrocortisone or equivalent), the relative risk for mortality at 28 days was 0.80 (95% CI: 0.67 to 0.95).
Qualitative signal: **Opposing findings:** The original systematic review found no mortality benefit regardless of dose. A pivotal trial and subsequent meta-analysis showed definite reductions in mortality with low dose regimens given for at least 5 days.

Quantitative signal: **Change in effect magnitude of 50% or more:** the absolute risk reduction for mortality increased from 0.2% to 4% (the criterion was first met at after Slusher 1996, when updated risk reduction increased to 1%)

Other signals: Increase in number of patients of at least 50%: N=530 → N=1067
Increase in number of trials of at least 50%: 10 trials → 16 trials

Change in width of 95% confidence interval of at least 50%: The original 95% CI for mortality with low-dose steroids extended from a 20% absolute reduction to a 16% increase in mortality. The 95% CI for the updated result extended from a 13% reduction to a 1% increase.

Source(s) of new evidence: **Pivotal trial:**

Additional trials and meta-analysis:

Time to signal: Qualitative signal: 7.1 years
Quantitative signal: 1 year


Question(s) addressed: Does metformin improve pregnancy and ovulation rates in women with polycystic ovary syndrome?
Findings of Original Review
The original review found that “metformin is effective in achieving ovulation in women with polycystic ovary syndrome, with odds ratios of 3.88 (95% confidence interval 2.25 to 6.69) for metformin compared with placebo and 4.41 (2.37 to 8.22) for metformin and clomifene compared with clomifene alone. An analysis of pregnancy rates shows a significant treatment effect for metformin and clomifene (odds ratio 4.40, 1.96 to 9.85).” Referring to the use of metformin, the authors concluded “its choice as a first line agent seems justified.”

New Findings
A pivotal trial compared clomifene citrate plus metformin with clomifene plus placebo and found a lower ovulation rate in the metformin group “(64% compared with 72% in the placebo group, a non-significant difference (risk difference − 8%, 95% confidence interval − 20% to 4%). There were no significant differences in either rate of ongoing pregnancy (40% v 46%; − 6%, − 20% to 7%) or rate of spontaneous abortion (12% v 11%; 1%, − 7% to 10%). A significantly larger proportion of women in the metformin group discontinued treatment because of side effects (16% v 5%; 11%, 5% to 16%).” The authors concluded that “metformin is not an effective addition to clomifene citrate as the primary method of inducing ovulation in women with polycystic ovary syndrome." The accompanying editorial also concluded that “metformin should not be used routinely as part of first line treatment for inducing ovulation.”

Qualitative signal
Opposing findings: The original systematic review concluded that metformin is definitely effective, recommending it as a first line agent. A pivotal trial showed no benefit and concluded that metformin should not be considered a first line treatment.

Quantitative signal
Change in statistical significance: increase in ovulation rate in patients treated with metformin and clomifene vs. clomifene alone lost statistical significance

Odds ratio of 4.41 (2.37, 8.22) → 1.42 (0.98, 2.05)

Change in effect magnitude of 50% or more:
Relative increase in ovulation rate in patients treated with metformin and clomifene vs. clomifene alone decreased by over 50% (OR of 4.41→ 1.42), as did the relative increase in clinical pregnancy rate among patients who received metformin and clomifene vs. clomifene alone (OR of 4.40→ 2.07)

Other signals:
Increase in number of patients of at least 50%: for the outcome of clinical pregnancy rate, the number of patients increased from 173 to 537

Increase in number of trials of at least 50%: for the outcome of clinical pregnancy rate, the number of trials increased from 3 to 8
Change in width of 95% confidence interval of at least 50%: as shown above

**Source(s) of new evidence**

**Pivotal trial:**

Four additional trials contained in meta-analysis:


**Time to signal**

Qualitative signal: 2.6 years
Quantitative signal: same

---

### II. Examples of Reviews with Signals for “Major Changes in Evidence” (criteria A4-A7)

**Examples of criterion A4:** Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook does not contradict the previous review, but characterizes benefit in substantially different terms (e.g., therapy previously characterized as “promising”, “likely beneficial” or similar description and now characterized as definitely beneficial.)

**Original Review**


**Question(s) addressed**

Covered a variety of questions related to the effects of antiplatelet therapy among patients at high risk of occlusive vascular events, including: Is aspirin plus dipyridamole was more effective than aspirin alone for the secondary prevention of vascular events after ischemic stroke of presumed arterial origin?

**Findings of Original Review**

The original review stated that “the addition of dipyridamole to aspirin was associated with only a non-significant further 6% (6%) reduction in serious vascular events…The apparent reduction in non-fatal stroke was derived mainly from one large study… but this result was not supported by the findings for non-fatal stroke in the other studies.” It concluded: “Addition of dipyridamole to aspirin produced no significant further reduction in vascular events compared with aspirin alone.”
New Findings

A pivotal trial found that patients who received aspirin and dipyridamole had a significantly lower risk of the primary outcome (a composite of death from all vascular causes, non-fatal stroke, non-fatal myocardial infarction, or major bleeding complication, whichever happened first), with a hazard ratio 0.80, 95% CI 0.66-0.98; absolute risk reduction 1% per year, 95% CI 0.1-1.8). Combining these data with previous trials resulted in an overall risk ratio for the composite of vascular death, stroke, or myocardial infarction of 0.82 (95% CI 0.74-0.91). The authors concluded: “The ESPRIT results, combined with the results of previous trials, provide sufficient evidence to prefer the combination regimen of aspirin plus dipyridamole over aspirin alone as antithrombotic therapy after cerebral ischaemia of arterial origin.”

Qualitative signal

Major change: possibly superior → definitely superior

Quantitative signal

Change in statistical significance: The lower risk of serious vascular events (vascular death or death from unknown cause, MI or stroke) became statistically significant.

Odds ratio of 0.94 (0.84, 1.06) → 0.90 (0.81, 0.99)

As noted above, the random effects meta-analytic result for relative risk is 0.82 (0.74-0.91), which more clearly shows the change in statistical significance. Odds ratios were used in our analysis because the original review used odds ratios.

Other signals

Because the original review covered a number of distinct questions related to antiplatelet therapy for the prevention of vascular events, other qualitative and quantitative and signals may have been met. For example, a pivotal trial found that adding aspirin to clopidogrel increased bleeding without reducing recurrent ischemic vascular events in high-risk patients. Another pivotal trial found that clopidogrel plus aspirin did not differ from aspirin alone for reducing MI, stroke, and cardiovascular death in patients with clinically evident cardiovascular disease or multiple risk factors.

Both of these qualitative signals occurred prior to the signal involving the comparison of aspirin plus dipyridamole with aspirin alone, but the latter more clearly fit one of our qualitative criteria and involved a quantitative signal as well.

Source(s) of Pivotal trial:

Additional pivotal trials addressing other questions in the original review:

Time to signal
Qualitative signal: 4.4 years
Quantitative signal: same

Original Review

Question(s) addressed
Do beta-blockers reduce mortality and morbidity in the treatment of heart failure?

Findings of Original Review
The original review reported a lower odds of death with beta-blockers that had borderline statistical significance (OR = 72; 99% CI 0.51 to 1.00). The authors were concerned about the sparseness of the data on mortality compared with evaluations of beta-blockers of patients with myocardial infarction. They concluded: “Although the effects on mortality were nominally statistically significant, the use of formal methods of interim monitoring adapted for meta-analyses suggests that substantially more patients still need to be studied in large scales trials to provide reliable and conclusive evidence.”

New Findings
A pivotal trial (MERIT-HF 1999) was stopped early because of the magnitude of reduction in the beta-blocker group, with a relative risk 0.66 (95% CI 0.53-0.81; p=0.0009 or adjusted for interim analyses p=0.0062). The authors concluded: “Metoprolol CR/XL once daily in addition to optimum standard therapy improved survival.” A second pivotal trial (CIBIS-II 1999) published the same year was also stopped early because of the survival benefit evident in the beta-blocker group. A third pivotal trial (Packer 2001) demonstrated a significant reduction in mortality for patients with more severe heart failure.

Qualitative Major change: possible mortality benefit → definite benefit signal
**Quantitative signal**

**Change in statistical significance:** borderline reduction in mortality became statistically significant

0.72 (0.51, 1.00) \(\rightarrow\) 0.67 (0.49, 0.91)

This change reflects the first shift to statistical significance (after Herlitz 1997); after additional trials, the updated result was 0.78 (0.70, 0.88)

**Other signals:**

Increase in number of trials of at least 50%: 10 trials \(\rightarrow\) 15 trials

Increase in number of patients of at least 50%: N= 2841 \(\rightarrow\) N=14738

Trial with sample size at least 3 times the size of previous largest trial: previous largest trial included 1094 patients; a new trial included 3991 patients

**Source(s) of evidence**

**Pivotal trial:**


**Additional trials (including two pivotal trials):**


**Time to signal**

Qualitative signal: 1 year
Quantitative signal: -0.6 years

**Original Review**


**Question(s) addressed**

Does prophylactic acetylcysteine reduces contrast nephropathy in patients with chronic renal insufficiency?

**Findings of Original Review**

The original review included 7 trials and found that “compared with periprocedural hydration alone, administration of acetylcysteine and hydration significantly reduced the relative risk of contrast nephropathy by 56% (0.435
[95% CI 0.215-0.879], p =0.02) in patients with chronic renal insufficiency. Meta-regression revealed no significant relation between the relative risk of contrast nephropathy and the volume of radiocontrast media administered or the degree of chronic renal insufficiency before the procedure.” The authors acknowledged that it remained unclear to what extent acetylcysteine improved harder clinical endpoints, but the impact on measures of renal function was regarded as robust. They concluded “acetylcysteine with hydration significantly reduces the risk of contrast nephropathy in patients with chronic renal insufficiency.”

**New Findings**

A subsequent meta-analysis (published 1.4 years after the first) included 20 trials and found that the impact on contrast nephropathy was smaller in magnitude and of borderline statistical significance. The authors also emphasized that the trials showed significant heterogeneity that remained unexplained despite exploration of various possible clinical and methodological differences across the studies.

They concluded: “Acetylcysteine may reduce the incidence of contrast-related nephropathy, but this finding is reported inconsistently across currently available trials. High-quality, large clinical trials are needed before acetylcysteine use in this indication can be recommended universally.”

**Qualitative signal**

**Major change:** Definite benefit → possible benefit

**Quantitative signal**

**Change in statistical significance:** the relative risk of contrast nephropathy with acetylcysteine versus hydration alone lost its statistical significance

RR of 0.44 (0.22, 0.88) → 0.81 (0.58, 1.13)

The loss of statistical significance first occurred with Gomes 2003, at which time the updated result was 0.61 (0.37, 1.00)

**Change in effect magnitude of 50% or more:** The relative risk reduction (RRR) decreased from 0.66 to 0.19

**Other signals:**

Increase in number of trials of at least 50%: 7 trials → 17 trials (20 trials included in newer meta-analysis, but not all provided data on the primary outcome)

Increase in number of patients of at least 50%: N=805 → N=1964

**Source(s) of Newer meta-analysis:**


**Time to signal** Qualitative signal: 1.4 years
Quantitative signal: -0.2 years

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Question(s) addressed</td>
<td>How do anti-leukotriene agents compare with inhaled glucocorticoids in terms of efficacy and safety in the management of chronic asthma?</td>
</tr>
<tr>
<td>Findings of Original Review</td>
<td>The original review showed non significant trends towards superiority of inhaled corticosteroids, but found the evidence insufficient to permit reliable conclusions regarding relative efficacy of the two treatments. The reviewers concluded: “Anti-leukotriene agents had a similar rate of exacerbations compared to inhaled corticosteroids, but inhaled steroids produced better lung function and quality of life as well as reduced symptoms, night awakenings and need for rescue beta2-agonist. Reliable conclusions cannot yet be drawn regarding the efficacy of this treatment due to the paucity of trials published in full text.”</td>
</tr>
<tr>
<td>New Findings</td>
<td>A subsequent update of the original review reported: “Patients treated with anti-leukotrienes were 60% more likely to suffer an exacerbation requiring systemic steroids…Significant differences favouring ICS were noted in most secondary outcomes, eg improvement in FEV1…symptom scores… Other significant benefits of ICS were seen for nocturnal awakenings, rescue medication use, and quality of life. Risk of side effects was not different between groups, but anti-leukotriene therapy was associated with 30% increased risk of &quot;withdrawals for any cause&quot; or &quot;withdrawals due to poor asthma control&quot;. The updated review concluded, “For most asthma outcomes, ICS at 400 mcg/day of beclomethasone-equivalent are more effective than anti-leukotriene agents given in the usual licensed doses... Inhaled glucocorticoids should remain the first line monotherapy for persistent asthma.”</td>
</tr>
<tr>
<td>Qualitative signal</td>
<td><strong>Major change:</strong> possibly inferior → definitely inferior</td>
</tr>
</tbody>
</table>
Quantitative signal: **Change in statistical significance**: The risk of asthma exacerbations with anti-leukotrienes vs inhaled steroids (in adults and children) became statistically significant.

Relative risk of 1.34 (0.93, 1.91) → 1.45 (1.07, 1.97)

**Other signals**: Increase in number of patients of at least 50%: N=1050 → N=1938

Increase in number of trials of at least 50%: 4 trials → 6 trials (for the above outcome)

**Source(s) of Evidence**


**New trials included in the meta-analysis**


**Time to signal**

Qualitative signal: 2 years

Quantitative signal: 0.4 years (became positive with Bleecker 2000)

**Example of criterion A5 for ‘Expansion of treatment’:** Pivotal trial, new or discordant meta-analysis, trial indexed in *ACP J Club*, more recent practice guideline, or recent textbook has expanded the role of the treatment (e.g., the treatment has now been shown to be of benefit in children or the elderly; or benefit now shown to apply to primary prevention of disease, not just secondary prevention).

**Original Review**


**Question(s)**

Does dexamethasone administered as an adjunct to antibiotic therapy improve
addressed outcomes for patients with bacterial meningitis, and does effectiveness vary by subcategories of causative organisms and timing or nature of antibiotic therapy?

Findings of Original Review
The original review found that “in Haemophilus influenzae type b meningitis, dexamethasone reduced severe hearing loss overall (combined odds ratio [OR], 0.31; 95% confidence interval [CI], 0.14-0.69)” and “in pneumococcal meningitis, only studies in which dexamethasone was given early suggested protection, which was significant for severe hearing loss (combined OR, 0.09; 95% CI, 0.0-0.71) and approached significance for any neurological or hearing deficit (combined OR, 0.23; 95% CI, 0.04-1.05).” The authors concluded that “The available evidence on adjunctive dexamethasone therapy confirms benefit for H influenzae type b meningitis and, if commenced with or before parenteral antibiotics, suggests benefit for pneumococcal meningitis in childhood.” The review contained only one study that included some adults (up to age 25 years of age).

New Findings
A pivotal trial that focused on adults patients and administered dexamethasone before or with the first dose of antibiotic and was given every 6 hours for four days showed: “treatment with dexamethasone was associated with a reduction in the risk of an unfavorable outcome (relative risk, 0.59; 95 percent confidence interval, 0.37 to 0.94; p =0.03). Treatment with dexamethasone was also associated with a reduction in mortality (relative risk of death, 0.48; 95 percent confidence interval, 0.24 to 0.96; p =0.04). Among the patients with pneumococcal meningitis, there were unfavorable outcomes in 26 percent of the dexamethasone group, as compared with 52 percent of the placebo group (relative risk, 0.50; 95 percent confidence interval, 0.30 to 0.83; p =0.006).” The authors concluded that “early treatment with dexamethasone improves the outcome in adults with acute bacterial meningitis and does not increase the risk of gastrointestinal bleeding.”

Qualitative signal
Major change: benefit reported in original review expanded to a new patient population
The original review concluded adjunctive dexamethasone conferred benefit only in children with acute bacterial meningitis due to Haemophilus influenzae type b and possibly pneumococcal meningitis. A pivotal trial showed significant benefit for adjunctive dexamethasone in adults with acute bacterial meningitis.

Quantitative signal
Not applicable

Other signals:
None

Source(s) of Pivotal trial:

Time to signal  Qualitative signal: 5.2 years
Quantitative signal: Not applicable

Example of criterion A6 for Important caveat: Pivotal trial, new meta-analysis, more recent practice guideline, or recent textbook adds an important caveat, about the patient populations who benefit, way in which treatment has to be delivered in order to derive benefit, sustainability of benefit (e.g., benefits on short term outcomes, but not long-term ones), or increases in harm that are not sufficient to undermine use altogether, but would clearly affect the decision to recommend treatment for at least some patient populations.


Question(s) addressed  How efficacious is allergen immunotherapy in controlling the symptoms, improving lung function, or decreasing the requirements for medication use in patients with asthma?

Findings of Original Review  The original review included 20 randomized placebo controlled double-blind trials and reported that “combined odds of symptomatic improvement from immunotherapy with any allergen were 3.2 (95% CI 2.2 to 4.9). The odds for reduction in medication after mite immunotherapy were 4.2 (95% CI 2.2 to 7.9). The combined odds for reduction in BHR [bronchial hyperreactivity] were 6.8 (95% CI 3.8 to 12.0). The mean effect size for any allergen immunotherapy on all continuous outcomes was 0.71 (95% CI 0.43 to 1.00), which would correspond to a mean 7.1% predicted improvement in FEV1 from immunotherapy.”

The authors also pointed out that “Although the benefits of allergen immunotherapy could be overestimated because of unpublished negative studies, an additional 33 such studies would be necessary to overturn these results.” They thus concluded that “allergen immunotherapy is a treatment option in highly selected patients with extrinsic ("allergic") asthma.”

New Findings  A pivotal trial reported that: “During the two treatment years, the mean peak expiratory flow rate was higher in the immunotherapy group (489 +/- 16 liters per minute, vs. 453 +/- 17 in the placebo group [p =0.06] during the first year, and 480 +/- 12 liters per minute, vs. 461 +/- 13 in the placebo group [p =0.03] during the second). Medication use was higher in the
immunotherapy group than in the placebo group during observation and lower during the first treatment year (p =0.01) but did not differ in the two groups during the second year (p =0.7). Asthma-symptom scores were similar in the two groups (p =0.08 in year 1 and p =0.3 in year 2). The immunotherapy group had reduced hay-fever symptoms, skin-test sensitivity to ragweed, and sensitivity to bronchial challenges and increased IgG antibodies to ragweed as compared with the placebo group; there was no longer a seasonal increase in IgE antibodies to ragweed allergen in the immunotherapy group after two years of treatment. Reduced medication costs were counterbalanced by the costs of immunotherapy.”

The authors concluded that “Although immunotherapy for adults with asthma exacerbated by seasonal ragweed exposure had positive effects on objective measures of asthma and allergy, the clinical effects were limited and many were not sustained for two years.”

<table>
<thead>
<tr>
<th>Qualitative signal</th>
<th>Major change: important caveat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In this case, the caveat concerns the sustainability of benefit.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantitative signal</th>
<th>None met</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Other signals:</th>
<th>None</th>
</tr>
</thead>
</table>

|---------------------------|----------------------------------------------------------------------------------------------------------------------------------|

<table>
<thead>
<tr>
<th>Time to signal</th>
<th>Qualitative signal: 327 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quantitative signal: Not applicable</td>
</tr>
</tbody>
</table>

**Example of criterion A7 for Opposing findings from discordant meta-analysis or non-pivotal trial:** The treatment has been characterized in sufficiently different terms to the cohort review that disagreement would have met criteria for ‘potentially invalidating change’ (A1) except the source was not a pivotal trial, new-meta-analysis, or more recent practice guideline, or recent textbook—rather, it was a discordant meta-analysis or trial indexed in ACP J Club.

|-----------------|----------------------------------------------------------------------------------------------------------------------------------|

<table>
<thead>
<tr>
<th>Question(s) addressed</th>
<th>Does pentoxifylline improve the walking capacity of patients with moderate intermittent claudication?</th>
</tr>
</thead>
</table>

| Findings of | The original meta-analysis found “a statistically significant improvement in |
the pain-free walking distance after pentoxifylline therapy (weighted mean difference 29.4 m [95% confidence interval (CI) 13.0 to 45.9 m])... A significant improvement was also noted in the absolute claudication distance (weighted mean difference 48.4 m [95% CI 18.3 to 78.6 m]). The authors concluded that “pentoxifylline therapy may be efficacious in improving the walking capacity of patients with moderate intermittent claudication.”

A randomized trial with a commentary in *ACP Journal Club* (Dawson 2002) compared pentoxifylline with an alternative medication, cilostazol, and placebo. The authors reported: “Mean maximal walking distance of cilostazol-treated patients (n = 227) was significantly greater at every postbaseline visit compared with patients who received pentoxifylline (n = 232) or placebo (n = 239). After 24 weeks of treatment, mean maximal walking distance increased by a mean of 107 m (a mean percent increase of 54% from baseline) in the cilostazol group, significantly more than the 64-m improvement (a 30% mean percent increase) with pentoxifylline (p <0.001). The improvement with pentoxifylline was similar (p =0.82) to that in the placebo group (65 m, a 34% mean percent increase).”

The authors concluded that “Cilostazol was significantly better than pentoxifylline or placebo for increasing walking distances in patients with intermittent claudication... Pentoxifylline and placebo had similar effects.”

The seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy (Clagett 2004) and *UpToDate* characterize pentoxifylline as no better than exercise and quote the above trial as the basis for this assessment.

**Qualitative signal**

**Major change**: possibly beneficial → definitely not beneficial

The original review concluded that pentoxifylline was likely efficacious in the treatment of intermittent claudication. A major practice guideline and chapter in recent textbook characterize pentoxifylline as no better than placebo based on the results of a trial that did not meet criteria for pivotal but was indexed in *ACP Journal Club*.

**Quantitative signal**

None met

**Other signals**: None

**Source(s) of new evidence**

Trial indexed in ACP J Club:


Practice guideline:

Clagett GP, Sobel M, Jackson MR, Lip GY, Tangelder M, Verhaeghe R. Antithrombotic therapy in peripheral arterial occlusive disease: the Seventh
ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004;126(3 Suppl):609S-626S.

**Time to signal** Qualitative signal: 4.1 years
Quantitative signal: not applicable
Appendix C: Sample Subject Searches

Examples of subject searches and how search features are scored. The last line of each search retrieved the Clinical Query set. The second from last line retrieves the Core Clinical Journal RCT set. The third from last line retrieves the meta-analysis set.


Feature scoring: 5 terms and exploded terms (lines 1,2, 3). The clinical query used is the sensitivity query.
1. exp Heparin/
2. exp Warfarin/
3. exp Venous Thrombosis/
4. Arthroplasty, Replacement, Knee/
5. Arthroplasty, Replacement, Hip/
6. 1 or 2
7. 4 or 5
8. 3 and 6 and 7
9. limit 8 to (yr="2000 - 2006" and meta analysis)
10. limit 8 to ("core clinical journals (aim)" and yr="2000 - 2006" and randomized controlled trial)
11. limit 8 to ("therapy (specificity)" and yr="2000 - 2006")


Feature scoring: 4 terms, exploded terms (line 1) and free text terms (line 3)
1. exp Indomethacin/
2. Ductus Arteriosus, Patent/
3. Cerebral Hemorrhage/ or intraventricular hemorrhage.mp.
4. 2 or 3
5. 1 and 4
6. limit 5 to (yr="1994 - 2006" and meta analysis)
7. limit 5 to ("core clinical journals (aim)" and yr="1994 - 2006" and randomized controlled trial)
8. limit 5 to ("therapy (optimized )" and yr="1994 - 2006")


Feature scoring: 2 terms, starred terms (lines 1 and 2) and subheadings (line 2).
1. *Angioplasty, Transluminal, Percutaneous Coronary/
2. *Myocardial Infarction/tu, dt
3. 1 and 2
4. limit 3 to (yr="1995 - 2006" and meta analysis)
5. limit 3 to ("core clinical journals (aim)" and yr="1995 - 2006" and randomized controlled trial)
6. limit 3 to ("therapy (optimized )" and yr="1995 - 2006")
Appendix D: Sample Assessment Worksheet

From Cohort ID 167


<table>
<thead>
<tr>
<th>Cohort Refid</th>
<th>Study refid</th>
<th>Author</th>
<th>Pub Year</th>
<th>MHDA</th>
<th>PMID</th>
<th>Type</th>
<th>On topic?</th>
<th>Eligible N</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>167</td>
<td>1</td>
<td>Mahony</td>
<td>1985</td>
<td>01/05/1985</td>
<td>3998921</td>
<td>Original</td>
<td>104</td>
<td>N for original is taken from Figure 2 unless otherwise noted</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>2</td>
<td>Ment</td>
<td>1985</td>
<td>01/12/1985</td>
<td>3906073</td>
<td>Original</td>
<td>48</td>
<td>N original is taken from Figure 2 unless otherwise noted</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>3</td>
<td>Vincen</td>
<td>1987</td>
<td>01/11/1987</td>
<td>3321891</td>
<td>Original</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>4</td>
<td>Rennie</td>
<td>1986</td>
<td>01/03/1986</td>
<td>3516077</td>
<td>Original</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>5</td>
<td>Krueger</td>
<td>1987</td>
<td>01/11/1987</td>
<td>3312552</td>
<td>Original</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>6</td>
<td>Hanigan</td>
<td>1988</td>
<td>01/06/1988</td>
<td>3373404</td>
<td>Original</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>7</td>
<td>Ment</td>
<td>1988</td>
<td>01/06/1988</td>
<td>3373405</td>
<td>Original</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>8</td>
<td>Bandstra</td>
<td>1988</td>
<td>01/10/1988</td>
<td>3174314</td>
<td>Original</td>
<td>199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>9</td>
<td>Bada</td>
<td>1989</td>
<td>01/10/1989</td>
<td>2677294</td>
<td>Original</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>10</td>
<td>Ment</td>
<td>1994</td>
<td>01/06/1994</td>
<td>8201485</td>
<td>Original</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>11</td>
<td>Ment</td>
<td>1994</td>
<td>01/04/1994</td>
<td>8124206</td>
<td>Original</td>
<td>431</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>12</td>
<td>Setzer</td>
<td>1984</td>
<td></td>
<td></td>
<td>MEDLINE</td>
<td>59</td>
<td>from table 2 of refid 167</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>13</td>
<td>Puckett</td>
<td>1985</td>
<td></td>
<td></td>
<td>MEDLINE</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>14</td>
<td>Bandstra</td>
<td>1987</td>
<td></td>
<td></td>
<td>MEDLINE</td>
<td>199</td>
<td>from table 2</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>76</td>
<td>Ohlsson</td>
<td>2005</td>
<td>24/02/2006</td>
<td>16235321</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>77</td>
<td>Loe</td>
<td>2005</td>
<td>11/08/2005</td>
<td>15994634</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>78</td>
<td>Thomas</td>
<td>2005</td>
<td>21/07/2005</td>
<td>15717178</td>
<td>MA</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>79</td>
<td>Simmer</td>
<td>2005</td>
<td>19/07/2005</td>
<td>15846747</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>80</td>
<td>Stevens</td>
<td>2004</td>
<td>30/11/2004</td>
<td>15266470</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>81</td>
<td>Herrera</td>
<td>2004</td>
<td>29/06/2004</td>
<td>14974018</td>
<td>MA</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>82</td>
<td>Ohlsson</td>
<td>2004</td>
<td>29/06/2004</td>
<td>14973955</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>83</td>
<td>Fowlie</td>
<td>2003</td>
<td>10/12/2003</td>
<td>14602691</td>
<td>MA</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>84</td>
<td>Cooke</td>
<td>2003</td>
<td>28/07/2003</td>
<td>12804488</td>
<td>MA</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>85</td>
<td>Shah</td>
<td>2003</td>
<td>27/03/2003</td>
<td>12535425</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>86</td>
<td>Brion</td>
<td>1999</td>
<td>26/07/1999</td>
<td>10353408</td>
<td>MA</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>87</td>
<td>Rubino</td>
<td>1997</td>
<td>05/06/1997</td>
<td>9165939</td>
<td>MA</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>88</td>
<td>Clark</td>
<td>1996</td>
<td>03/01/1997</td>
<td>8951253</td>
<td>MA</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>89</td>
<td>Clyman</td>
<td>1996</td>
<td>26/06/1996</td>
<td>8627430</td>
<td>MA</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>74</td>
<td>Simko</td>
<td>1994</td>
<td>01/06/1994</td>
<td>8169675</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>72</td>
<td>Rush</td>
<td>1994</td>
<td>29/09/1994</td>
<td>8071758</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>71</td>
<td>Hammerma n</td>
<td>1995</td>
<td>27/02/1995</td>
<td>7838642</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>70</td>
<td>Ment</td>
<td>1995</td>
<td>14/04/1995</td>
<td>7892866</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>69</td>
<td>Corbet</td>
<td>1995</td>
<td>07/07/1995</td>
<td>7776110</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>68</td>
<td>Malloy</td>
<td>1995</td>
<td>11/10/1995</td>
<td>7666265</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>66</td>
<td>Bernstein</td>
<td>1996</td>
<td>07/06/1996</td>
<td>8618177</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>67</td>
<td>Van</td>
<td>1995</td>
<td>07/06/1996</td>
<td>8618789</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>65</td>
<td>Couser</td>
<td>1996</td>
<td>26/06/1996</td>
<td>8627434</td>
<td>Candidate</td>
<td>Y</td>
<td>Y</td>
<td>90 Identified in ref id 83 and included in its</td>
</tr>
</tbody>
</table>

New comparison b/w prolonged and short course of Indomethacin without definitive results
<table>
<thead>
<tr>
<th>Cohort Refid</th>
<th>Study refid</th>
<th>Author</th>
<th>Pub Year</th>
<th>MHDA</th>
<th>PMID</th>
<th>Type</th>
<th>On topic?</th>
<th>Eligible N</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>167</td>
<td>64</td>
<td>Ment</td>
<td>1996</td>
<td>01/07/1996</td>
<td>8632956</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>63</td>
<td>Silver</td>
<td>1996</td>
<td>15/08/1996</td>
<td>8677067</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>62</td>
<td>Rastogi</td>
<td>1996</td>
<td>29/08/1996</td>
<td>8692619</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>61</td>
<td>Ment</td>
<td>1996</td>
<td>04/12/1996</td>
<td>8885951</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>60</td>
<td>Gerstmann</td>
<td>1996</td>
<td>03/01/1997</td>
<td>8951252</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>59</td>
<td>Yaseen</td>
<td>1997</td>
<td>12/05/1997</td>
<td>9078828</td>
<td>Candidate</td>
<td>Y</td>
<td>Y</td>
<td>27 Identified in ref id 83 and included in its pooled stats</td>
</tr>
<tr>
<td>167</td>
<td>58</td>
<td>Van</td>
<td>1997</td>
<td>26/06/1997</td>
<td>9175948</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>57</td>
<td>Allan</td>
<td>1997</td>
<td>10/07/1997</td>
<td>9193243</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>56</td>
<td>Hudak</td>
<td>1997</td>
<td>21/07/1997</td>
<td>9200358</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>55</td>
<td>Parilla</td>
<td>1997</td>
<td>04/08/1997</td>
<td>9215169</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>54</td>
<td>Romagnoli</td>
<td>1997</td>
<td>11/09/1997</td>
<td>9284854</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>53</td>
<td>Keszler</td>
<td>1997</td>
<td>14/10/1997</td>
<td>9310511</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>52</td>
<td>Chugh</td>
<td>1997</td>
<td>13/11/1997</td>
<td>9332109</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>50</td>
<td>Su</td>
<td>1998</td>
<td>30/04/1998</td>
<td>9553290</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>49</td>
<td>Morales</td>
<td>1998</td>
<td>21/05/1998</td>
<td>9568218</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>48</td>
<td>Ment</td>
<td>1998</td>
<td>14/08/1998</td>
<td>9714645</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>51</td>
<td>Yeh</td>
<td>1997</td>
<td>10/09/1998</td>
<td>9310536</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>46</td>
<td>Anand</td>
<td>1999</td>
<td>16/04/1999</td>
<td>10201714</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>45</td>
<td>Tammela</td>
<td>1999</td>
<td>25/05/1999</td>
<td>10228288</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>44</td>
<td>Kothadia</td>
<td>1999</td>
<td>22/07/1999</td>
<td>10390255</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>43</td>
<td>Vohr</td>
<td>1999</td>
<td>26/08/1999</td>
<td>10405190</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>42</td>
<td>Panter</td>
<td>1999</td>
<td>02/09/1999</td>
<td>10430197</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>41</td>
<td>Kopelman</td>
<td>1999</td>
<td>12/10/1999</td>
<td>10484801</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>40</td>
<td>Baenziger</td>
<td>1999</td>
<td>11/01/2000</td>
<td>10592922</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>36</td>
<td>Patel</td>
<td>2000</td>
<td>20/01/2000</td>
<td>10625080</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>35</td>
<td>Ment</td>
<td>2000</td>
<td>22/03/2000</td>
<td>10699097</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>39</td>
<td>Su</td>
<td>1999</td>
<td>23/03/2000</td>
<td>10525023</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>38</td>
<td>Supapannac hart</td>
<td>1999</td>
<td>05/04/2000</td>
<td>10730525</td>
<td>Candidate</td>
<td>Y</td>
<td>Y</td>
<td>30 Identified in ref id 83 and included in its pooled stats</td>
</tr>
<tr>
<td>167</td>
<td>34</td>
<td>Couser</td>
<td>2000</td>
<td>21/06/2000</td>
<td>10850507</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td>Outcome not considered</td>
</tr>
<tr>
<td>167</td>
<td>33</td>
<td>Van</td>
<td>2000</td>
<td>07/09/2000</td>
<td>10974130</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>32</td>
<td>De Carolis</td>
<td>2000</td>
<td>14/09/2000</td>
<td>10834523</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>37</td>
<td>Sanghvi</td>
<td>1999</td>
<td>05/01/2001</td>
<td>10740301</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>30</td>
<td>Stark</td>
<td>2001</td>
<td>11/01/2001</td>
<td>11150359</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>29</td>
<td>Wardle</td>
<td>2001</td>
<td>22/02/2001</td>
<td>11124916</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>28</td>
<td>Van</td>
<td>2001</td>
<td>05/04/2001</td>
<td>11174617</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>31</td>
<td>Dani</td>
<td>2000</td>
<td>31/05/2001</td>
<td>11106052</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>27</td>
<td>Schmidt</td>
<td>2001</td>
<td>05/07/2001</td>
<td>11430325</td>
<td>Candidate</td>
<td>Y</td>
<td>Y</td>
<td>1202 Identified in ref id 83 and included in its pooled stats</td>
</tr>
<tr>
<td>167</td>
<td>26</td>
<td>Vermont</td>
<td>2001</td>
<td>18/10/2001</td>
<td>11553345</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>25</td>
<td>Schmidt</td>
<td>2002</td>
<td>07/10/2002</td>
<td>12241754</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>24</td>
<td>Christmann</td>
<td>2002</td>
<td>22/11/2002</td>
<td>12061361</td>
<td>Candidate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>23</td>
<td>Lago</td>
<td>2002</td>
<td>04/02/2003</td>
<td>12014386</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>22</td>
<td>Supapannac hart</td>
<td>2002</td>
<td>12/02/2003</td>
<td>12549803</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>21</td>
<td>Vohr</td>
<td>2003</td>
<td>18/04/2003</td>
<td>12671149</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Cohort Ref</td>
<td>Study refid</td>
<td>Author</td>
<td>Pub Year</td>
<td>MHDA</td>
<td>PMID</td>
<td>Type</td>
<td>On topic?</td>
<td>Eligible N</td>
<td>Notes</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td>---------------</td>
<td>-----------</td>
<td>------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>167</td>
<td>20</td>
<td>Lee</td>
<td>2003</td>
<td>03/10/2003</td>
<td>12897285</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>19</td>
<td>Simons</td>
<td>2003</td>
<td>25/11/2003</td>
<td>14612478</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>18</td>
<td>Schreiber</td>
<td>2003</td>
<td>08/12/2003</td>
<td>14645637</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>17</td>
<td>Osborn</td>
<td>2003</td>
<td>10/12/2003</td>
<td>14602694</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>16</td>
<td>Chotigeat</td>
<td>2003</td>
<td>05/02/2004</td>
<td>14700149</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>15</td>
<td>Su</td>
<td>2003</td>
<td>26/02/2004</td>
<td>14651538</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>14</td>
<td>Kumar</td>
<td>2004</td>
<td>05/10/2004</td>
<td>15235161</td>
<td>Candidate</td>
<td>Y</td>
<td>Y</td>
<td>115</td>
</tr>
<tr>
<td>167</td>
<td>13</td>
<td>Gournay</td>
<td>2004</td>
<td>15/12/2004</td>
<td>15567009</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>12</td>
<td>Van</td>
<td>2004</td>
<td>15/12/2004</td>
<td>15567010</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>11</td>
<td>Vila-Vazqu</td>
<td>2004</td>
<td>18/03/2005</td>
<td>15298714</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>10</td>
<td>Mestan</td>
<td>2005</td>
<td>12/07/2005</td>
<td>16000353</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>9</td>
<td>Van Meurs</td>
<td>2005</td>
<td>12/07/2005</td>
<td>16000352</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>8</td>
<td>Hall</td>
<td>2005</td>
<td>27/09/2005</td>
<td>15867047</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>6</td>
<td>Bell</td>
<td>2005</td>
<td>10/11/2005</td>
<td>15930233</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>7</td>
<td>Dani</td>
<td>2005</td>
<td>10/11/2005</td>
<td>15930213</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>5</td>
<td>Gimeno</td>
<td>2005</td>
<td>07/12/2005</td>
<td>16219273</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>4</td>
<td>Olney</td>
<td>2005</td>
<td>29/12/2005</td>
<td>16041635</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>3</td>
<td>Zanardo</td>
<td>2005</td>
<td>19/01/2006</td>
<td>16260891</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>1</td>
<td>Mercer</td>
<td>2006</td>
<td>10/05/2006</td>
<td>16585320</td>
<td>Candidate</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>2</td>
<td>Adamska</td>
<td>2005</td>
<td>25/05/2006</td>
<td>16547381</td>
<td>Candidate</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td></td>
<td>Morales-Suarez</td>
<td>1994</td>
<td>MEDLINE n</td>
<td>Y Y 80</td>
<td>Identified in Cochrane version of Fowlie 2003. Analysis 01.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td></td>
<td>Domanico</td>
<td>1994</td>
<td>MEDLINE n</td>
<td>Y Y 100</td>
<td>Fowlie 2003 Analysis 01.04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note candidate studies after the last search date of ref id 83 were screened.
Appendix E: Searches by Clinical Area

Searches for each clinical area (a complete example, showing publication type and year limits is shown below.

Cardiac and cardiovascular systems
1. exp Heart Diseases/
2. Heart Failure, Congestive/
3. or/1-2

Neurology
1. exp Nervous System Diseases/

Gastroenterology and Hepatology
1. exp Gastrointestinal Diseases/

Critical Care Medicine
1. exp Critical Care/
2. exp Critical Illness/

Endocrinology and Metabolism
1. exp Bone Diseases, Metabolic/
2. exp Metabolic Diseases/
3. or/1-2

Infectious Diseases
1. exp Anti-infective agents/

Obstetrics and Gynecology
1. exp Genital Diseases, Female/
2. exp Pregnancy Complications/
3. or/1-2

Oncology
1. exp Neoplasms/

Peripheral Vascular Diseases
1. exp Vascular Diseases/

Psychiatry
1. exp Mental Disorders/

Respiratory System
1. exp Respiratory Tract Diseases/

Rheumatology
1. exp Rheumatic Diseases/

Urology and Nephrology
1. exp Kidney Diseases/

Example of a complete search:
Cardiac and cardiovascular systems
1. exp Heart Diseases/
2. Heart Failure, Congestive/
3. or/1-2
4. limit 3 to animals
5. remove duplicates from 4
6. limit 5 to yr="1988"
7. limit 5 to yr="1989"
8. limit 5 to yr="1990"
9. limit 5 to yr="1991"
10. limit 5 to yr="1992"
11. limit 5 to yr="1993"
12. limit 5 to yr="1994"
13. limit 5 to yr="1995"
14. limit 5 to yr="1996"
15. limit 5 to yr="1997"
16. limit 5 to yr="1998"
17. limit 5 to yr="1999"
18. limit 5 to yr="2000"
19. limit 5 to yr="2001"
20. limit 5 to yr="2002"
21. limit 5 to yr="2003"
22. limit 5 to yr="2004"
23. limit 5 to yr="2005"
24. limit 5 to yr="2006"
25. limit 5 to yr="2007"
26. limit 3 to randomized controlled trial
27. limit 3 to clinical trial
28. 27 not 26
29. remove duplicates from 28
30. limit 29 to yr="1988"
31. limit 29 to yr="1989"
32. limit 29 to yr="1990"
33. limit 29 to yr="1991"
34. limit 29 to yr="1992"
35. limit 29 to yr="1993"
36. limit 29 to yr="1994"
37. limit 29 to yr="1995"
38. limit 29 to yr="1996"
39. limit 29 to yr="1997"
40. limit 29 to yr="1998"
41. limit 29 to yr="1999"
42. limit 29 to yr="2000"
43. limit 29 to yr="2001"
44. limit 29 to yr="2002"
45. limit 29 to yr="2003"
46. limit 29 to yr="2004"
47. limit 29 to yr="2005"
48. limit 29 to yr="2006"
49. limit 29 to yr="2007"
50. limit 3 to randomized controlled trial
51. remove duplicates from 50
52. limit 51 to yr="1988"
53. limit 51 to yr="1989"
54. limit 51 to yr="1990"
55. limit 51 to yr="1991"
56. limit 51 to yr="1992"
57. limit 51 to yr="1993"
58. limit 51 to yr="1994"
59. limit 51 to yr="1995"
60. limit 51 to yr="1996"
61. limit 51 to yr="1997"
62. limit 51 to yr="1998"
63. limit 51 to yr="1999"
64. limit 51 to yr="2000"
65. limit 51 to yr="2001"
66. limit 51 to yr="2002"
67. limit 51 to yr="2003"
68. limit 51 to yr="2004"
69. limit 51 to yr="2005"
70. limit 51 to yr="2006"
71. limit 51 to yr="2007"
72. limit 3 to systematic reviews
73. remove duplicates from 72
74. limit 73 to yr="1988"
75. limit 73 to yr="1989"
76. limit 73 to yr="1990"
77. limit 73 to yr="1991"
78. limit 73 to yr="1992"
79. limit 73 to yr="1993"
80. limit 73 to yr="1994"
81. limit 73 to yr="1995"
82. limit 73 to yr="1996"
83. limit 73 to yr="1997"
84. limit 73 to yr="1998"
85. limit 73 to yr="1999"
86. limit 73 to yr="2000"
87. limit 73 to yr="2001"
88. limit 73 to yr="2002"
89. limit 73 to yr="2003"
90. limit 73 to yr="2004"
91. limit 73 to yr="2005"
92. limit 73 to yr="2006"
93. limit 73 to yr="2007"
94. limit 3 to guideline
95. remove duplicates from 94
96. limit 95 to yr="1988"
97. limit 95 to yr="1989"
98. limit 95 to yr="1990"
99. limit 95 to yr="1991"
100. limit 95 to yr="1992"
101. limit 95 to yr="1993"
102. limit 95 to yr="1994"
103. limit 95 to yr="1995"
104. limit 95 to yr="1996"
105. limit 95 to yr="1997"
106. limit 95 to yr="1998"
107. limit 95 to yr="1999"
108. limit 95 to yr="2000"
109. limit 95 to yr="2001"
110. limit 95 to yr="2002"
111. limit 95 to yr="2003"
112. limit 95 to yr="2004"
113. limit 95 to yr="2005"
114. limit 95 to yr="2006"
115. limit 95 to yr="2007"
Appendix F: Growth of the Literature by Clinical Area

Cumulative publications by type and clinical area
X axis represents number of records, Y axis is year in each case.
Appendix G: Survey Instrument

Wednesday, March 14, 2007

Updating Systematic Reviews Survey (University of Ottawa EPC)

Part I: Survey Information & Consent Form

Survey Title: Examining the Updating Practices of Healthcare Organizations that Fund and/or Conduct Systematic Review Research.

Background: To maintain their central role in informing clinical practice and health care policy, systematic reviews need to be up-to-date. However, there is little available guidance about the appropriate timing or approach for updating of a particular evidence-base. Furthermore, updating policies and practices of organizations are often not explicit. There are general concerns about dated reviews; however, unlike other methodological areas, methods for updating systematic reviews remain understudied. Invitation to Participate: Your organization/group/unit has been chosen to participate in the first phase of this international survey on systematic review updating experiences given its involvement as a producer and/or funder of individual systematic reviews, or systematic reviews that are incorporated into clinical practice guidelines (CPGs) or health technology assessments (HTAs). Benefits: To better understand the updating experiences of those engaged in systematic reviews research, your responses would facilitate the advancement of best practice of updating methodology. This research is also of relevance to evidence-based agencies and researchers wishing to increase general performance and efficiency in their respective updating practices.

This research is funded by the U.S. Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Program and is being undertaken by the University of Ottawa Evidence-based Practice Center (UO-EPC) as part of an ongoing research program to examine a broad spectrum of issues pertaining to the methodology for updating systematic reviews. This will also be part of my MSc thesis in Public Health Sciences, University of Toronto, Canada. The Principle Investigator of this study is Dr. David Moher, Director of the UO-EPC and the Chalmers Research Group, based at the Children’s Hospital of Eastern Ontario Research Institute (CHEO RI).

Instructions: As a participant, we are asking that you complete this survey, which is anticipated to take 30 minutes. Your organization’s participation is voluntary and at any time you can leave the survey by clicking on "Exit this survey". If you wish to return, your responses will be saved. You may withdraw from the study at any time and may refrain from answering any question by leaving it blank. All responses will remain strictly confidential and we anticipate minimal or no risk of harm from the research study. As a small token of our appreciation for completing this survey, you will be provided with a $10 gift certificate from Amazon.com. In order to obtain your consent, please review each statement below:

1. I am over 18 years of age and have the approval from my organization/group/unit to represent them as a participant of this survey;
2. I understand that the information obtained from the study will be used within a MSc thesis and that aggregate results from this survey may be published in peer-reviewed journals, however will not be directly identify organizations or groups by name in any reports or publications;
3. I understand that my organization’s participation is voluntary and withdrawal from the study is allowed at any stage without adverse consequences for my organization;
4. I understand in representing my organization, that I may refrain from answering any questions should I so wish;
5. I understand that apart from the $10 Amazon gift card being offered for completing this survey, that I may not directly benefit from taking part in the study and that I will not receive any payment for participating in this research;
6. I understand that my organization will be provided with an executive summary of the results upon request;
7. I understand the Children’s Hospital of Eastern Ontario and the University of Toronto’s Research Ethics Boards have formally approved this survey;
8. I understand that data collected for this research will be retained for a period of 3 years, and will be stored in a secure location;
9. I understand that all responses will be kept strictly confidential;
10. I understand that I waive no legal rights by participating in this survey.

If you have any questions related to this survey, please contact:

Chantelle Garrity
University of Ottawa EPC Coordinator
CHEO RI, Ottawa, Canada
Email: cgarrity@cheo.on.ca
Tel: 613-737-7600 x 4117
Fax: 613-738-4800

If you have any questions or concerns about your rights as a research participant, please contact Sharon Haig (shaig@cheo.on.ca).

On behalf of the UO-EPC, thank you in advance for your time.
* I AGREE TO TAKE PART IN THIS STUDY (Confirming my consent to participate)

Yes _______  No _______

On behalf of the University of Ottawa EPC, thank you for your response.

If you still wish to exit this survey, please click on the Exit This Survey link provided above in the upper right corner of the screen.

Otherwise, click on 'PREVIOUS' to return to the preceding page to agree to take part in this survey.

Please review the following before starting the survey.
Systematic Review refers to a study that has a clear and comprehensive methodology. It includes explicit research questions, methods of searching the literature, criteria for including material, criteria for appraising quality and reliability of studies, and synthesis of research findings.

Please Note: For the purposes of this survey, the term Systematic Review is intended to be used broadly to reflect the following:

i) Conduct of individual or stand alone systematic reviews or
ii) Conduct of systematic reviews for use in Clinical Practice Guidelines (CPGs), or Health Technology Assessments (HTAs)

Therefore, please complete this survey based on the overall research synthesis work with which your organization/group/unit may be involved either as a producer, funder and/or both.

1. Which of the following statements best describe your organization/group? (Please tick one of the following options)
   - A funder of systematic reviews (including as part of Clinical Practice Guidelines (CPGs), or Health Technology Assessments (HTAs))
   - A producer of systematic reviews (including as part of Clinical Practice Guidelines (CPGs), or Health Technology Assessments (HTAs))
   - Both a funder & producer of systematic reviews (including as part of Clinical Practice Guidelines (CPGs), or Health Technology Assessments (HTAs))

2. Which of the following statements best describe why your organization/group produces systematic reviews? (Please tick one of the following options)
   - For knowledge support (e.g., Cochrane reviews or other discrete systematic reviews)
   - For decision support (e.g., policy or clinical decision-making; evidence for HTA's or CPG's)
   - For both knowledge & decision support
   - Other (please specify)

3. We suggest UPDATE refers to "a discrete event aiming to search for and identify new evidence to incorporate into a previously completed systematic review." (New evidence is taken to mean any evidence not included in the previously completed review, irrespective of its' chronological appearance in the literature)

To what extent do you think your organization/group would agree or disagree with this definition? (Please tick one of the following options):

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree
- No opinion

4. Overall, what level of importance would you estimate your organization/group places on updating systematic reviews? (Please tick one of the following options)

- 1 (low)
- 2
- 3
- 4
- 5 (high)
5. How would you rank updating of systematic reviews compared to other research mandates within your organization/group? (Please tick one of the following options)

☐ 1 (low) 2 3 4 5 (high)

6. Does your organization/group have a policy on updating systematic reviews?

☐ Yes
☐ No
☐ Not sure

7. Is establishing a formal updating policy or process something your organization/group would view as important? (Please tick one of the following options)

☐ Yes
☐ No
☐ Not sure
☐ Not applicable

8. Please describe the updating policy of your organization/group (if applicable):
(or list any working documents detailing the updating policy of your organization in the box provided below)

9. In general, how would you describe the current ‘updating’ practices of your organization/group? (Please tick one of the following options)

☐ Update regularly
☐ Update irregularly
☐ Do not update

10. Of the systematic reviews either produced or commissioned by your organization/group, what percentage would you estimate are up to date?

☐ 0-10%
☐ 11-20%
☐ 21-30%
☐ 31-40%
☐ 41-50%
☐ 51-60%
☐ 61-70%
☐ 71-80%
☐ 81-90%
☐ 91-100%
☐ Not sure

11. Which of the following does your organization/group feel is the most responsible for ensuring systematic reviews are updated?

☐ Funder(s) of original review
☐ Authors of the original review
☐ Information specialist/ Librarian
☐ Policy-maker utilizing the evidence
☐ All of the above
12. How would you describe the method(s) used by your organization/group to determine the need to update a systematic review?

- Formal method(s) used
- Informal method(s) used
- No method(s) used
- Not sure

13. Does your organization/group conduct literature searches regularly to help identify new relevant literature?

- Yes
- No
- Not sure
- Not applicable

14. How often are searches conducted to identify new literature? Please tick one of the following options

- Monthly
- Every 6 months
- Every 12 months
- Every 18 months
- Not applicable
- Other time interval (please specify)

15. How often are the following search strategies typically used when monitoring the literature to identify new evidence? Please tick one answer for each type of strategy

<table>
<thead>
<tr>
<th></th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Same search strategy from the original review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Modified search strategy from the original</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- New search strategy - original search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16. How frequent does your organization/group use the following strategies to monitor emergence of new evidence for updating? (Please provide one answer for each strategy.)

<table>
<thead>
<tr>
<th></th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>- General literature searches (e.g. electronic, hand-search etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Trial registry surveillance (e.g., ClinicalTrials.Gov, WHO International Clinical Trials Registry, various national)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
government or industry trials registries)
- Statistical approaches (e.g., Barrowman's diagnostic test, cumulative meta-analysis etc.)
- Systematic reviews surveillance
Guideline or health technology assessment surveillance
- Experts in the field

17. Please list any additional strategies, actions, steps or methods used by your organization/group to monitor for new evidence,
(or list any working documents your organization has developed listing these monitoring strategies in the box provide below). If there are no additional strategies, actions, or steps that you are aware of, please leave blank.

18. How often do the following issues factor into your organization/group determining WHEN to conduct or fund an update? Please provide one answer for each reason

<table>
<thead>
<tr>
<th>Reason</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of new studies identified</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Number of participants in new studies</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Reporting of serious or 'new' serious adverse events</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>New inclusion criteria (outcomes; interventions; populations; methodological advances/new analysis)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Totality (comprehensiveness) of all new evidence or data (including harms &amp; Benefits)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Time credibility</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Need for an internal organizational decision</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Formal request from policy-maker or healthcare decision-maker</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

19. How often do each of the following individuals/groups impact your organization/group's decision-making process of whether to fund or conduct an update?
Please provide one answer for each reason

<table>
<thead>
<tr>
<th>Reason</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original authors of the systematic review (in-house or external to your organization)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Information specialist/Librarian</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Experts in the field</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Statistician(s)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>External Policy-maker(s)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
- The organization/group itself as original funder of a systematic review
- Patient(s) or consumer group

20. How often do the following issues impact your organization/group's decision-making process of whether to fund or conduct an update? Please tick one answer for each reason

<table>
<thead>
<tr>
<th>Reason</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost utility of updating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact on policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact on clinical practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current public controversy/interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burden of illness/costs of disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization/group credibility of being</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anticipated change in effect size or precision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To address a quality gap in care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

21. Please list any other additional issues that may factor into your organization/group determining WHEN to update a systematic review using the box provided below.
If there are no additional issues that you are aware of, please leave blank.

22. How would you best describe the methods or strategies used by your organization/group to determine HOW to conduct an update of a systematic review? Please tick one of the following options

- Formal method(s) used
- Informal method(s) used
- No method(s) used
- Not sure

23. Please describe the formal/informal method(s) in the box provided below (if applicable):

24. On average, which of the following best estimates the time your organization/group typically expends on implementing an updating strategy?
Please tick one of the following options

- 0-3 months
- 4-6 months
- 7-9 months
- 10-12 months
- 13-15 months
- 16-18 months
- 19-21 months
- 21-24 months
25. On average, what is your organization/group's estimated budget for implementing an updating strategy? (U.S. Dollars) Please tick one of the following options

- $<10,000
- $10,000 - $20,000
- $21,000 - $30,000
- $31,000 - $40,000
- $41,000 - $50,000
- $51,000 - $60,000
- $61,000 - $70,000
- $71,000 – $80,000
- $81,000 – $90,000
- $91,000 - $100,000
- $100,001 - $200,000
- $>200,000
- Not sure
- Not applicable

You have reached the half-way mark of this survey.

26. How often has your organization/group been involved in the following scenarios? Please tick one answer for each activity.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>A partial update involving only certain sections of a review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A full update of all sections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An entirely new review upon updating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowing a systematic review is out of date but unable to commence updating due to lack of resources (e.g. funding, personnel, time etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27. When you do an update, how often is your organization/group able to draw on the same people who did the original review? Please tick one of the following options

- Always
- Often
- Sometimes
- Seldom
- Never
- Not sure
28. How frequently does your organization/group use the following procedures when updating? Please tick one answer for each reason

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Time-specific approach (pre-set updating frequency)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Editorial strategy with an algorithm of administrative actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Bibliometric database entry-date searching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Cumulative meta-analysis (CMA) or its' extensions (e.g., recursive; with sequential boundaries; or using cumulative slope)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Barrowman's diagnostic test (i.e., identifying 'null' meta-analyses ready for updating)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

29. Please list any other additional procedures used during the updating process or list any relevant working documents your organization has developed detailing these procedures in the box provided below. If there are no new additional strategies used during the update process that you are aware of, please leave blank.

30. How often has your organization/group been involved in updating systematic reviews done by others? Please tick one of the following options

- Always ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Often ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Sometimes ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Seldom ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Never ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Not sure ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐
- Not applicable ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐  ☐

31. When updating reviews done by others, how often is your organization/group involved in the following: Please tick one answer for each activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Seldom</th>
<th>Never</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Revising the original search strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Revising the eligibility criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Re-abstracting data from previously reviewed studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Re-assessing quality of previously reviewed studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
32. To what extent are the following characteristics of original reviews perceived as barriers to updating reviews done by others? Please tick one answer for each reason

<table>
<thead>
<tr>
<th>Perceived need to change the original search strategy</th>
<th>Serious Barrier</th>
<th>Moderate Barrier</th>
<th>Minor Barrier</th>
<th>Not a Barrier</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived need to change the original screening criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived need to redo the data abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived need to reassess study quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

33. How often does your organization/group identify recent relevant literature published after the date of the last search but before completion of the final draft?

Please tick one of the following options

- Always
- Often
- Sometimes
- Seldom
- Never
- Not sure
- Not applicable

34. When recent relevant literature is identified as being published after the date of the last search but before completion of the final draft, how does your organization/group usually incorporate this new evidence?

Please check all that apply

- As an addendum in the review
- As a formal revision to the analysis
- Not sure
- Not applicable
- Other

35. To what extent would you consider the following as general barriers to your organization/group updating systematic reviews?

Please tick one answer for each reason

- Limited updating methodologies
- Limited funding/resources
- Work redundancy of updating (reviewer motivation)
- Limited academic credit for updating

<table>
<thead>
<tr>
<th>- Limited updating methodologies</th>
<th>Serious Barrier</th>
<th>Moderate Barrier</th>
<th>Minor Barrier</th>
<th>Not a Barrier</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Limited funding/resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Work redundancy of updating (reviewer motivation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Limited academic credit for updating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
work
- Limited journal publishing formats
- Having to update reviews done by others

36. To what extent does your organization/group agree or disagree with harmonizing updating efforts across organizations or groups that fund or conduct systematic reviews? Please tick one of the following options

☐ Strongly agree
☐ Somewhat agree
☐ Neither agree nor disagree
☐ Somewhat disagree
☐ Strongly disagree
☐ No opinion

37. To what extent would you consider the following as benefits to your organization/group participating in international harmonization efforts for updating systematic reviews? Please tick one answer for each reason

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Major benefit</th>
<th>Moderate benefit</th>
<th>Minor benefit</th>
<th>No benefit</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to new information, ideas, materials or other resources</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Potential to minimize duplication of services</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Use of existing resources more efficiently</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ability to address issues beyond a single organization/group's domain</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Shared responsibility across organization/group's for complex or controversial issues</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

38. To what extent would you consider the following as barriers to your organization/group participating in international harmonization efforts for updating systematic reviews? Please tick one answer for each reason

<table>
<thead>
<tr>
<th>Barrier</th>
<th>Serious Barrier</th>
<th>Moderate Barrier</th>
<th>Minor Barrier</th>
<th>Not a Barrier</th>
<th>Not sure</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversion of an organization/group's human resources</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Diversion of an organization/group's</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

G-11
funding resources
- Diversion from an organization/group's research mandate
- Perceived delays in working across organizations/groups

39. To what extent would your organization/group agree or disagree with development of a central registry of systematic reviews including existing systematic reviews and protocols (i.e., similar to efforts within the clinical trials community)? Please tick one of the following options

☐ Strongly agree
☐ Somewhat agree
☐ Neither agree nor disagree
☐ Somewhat disagree
☐ Strongly disagree
☐ No opinion

40. Does your organization/group encourage the following activities when funding or conducting systematic reviews? Please tick one answer for each activity

- Inclusion of a prediction for updating in the text of original reviews
- Withdrawal of systematic reviews from circulation when assessed as out of date
- Formal retirement of systematic reviews when deemed no longer in need of further investigation

41. Which of the following best describes your organization/group? Please tick one of the following options

☐ Academic Institution
☐ National Government Agency
☐ Regional/Local Government Agency
☐ Private Organization (Industry)
☐ Medical Specialty Society
☐ Disease Specific Society
☐ Managed Care Organization
☐ Other (please specify)

42. What is your organization/group's primary funding structure? Please tick one of the following options

☐ For profit
☐ Not for profit

43. What type(s) of funding does your organization receive? Please tick all that apply

☐ Industry/Private Sector
☐ Government (Infrastructure; grants)
☐ Non-profit (Academic; non-governmental organizations)
☐ Endowment fund
☐ Internal
☐ Other (please specify)

44. What is your organization/group's total estimated proportion of work expended annually on the following research synthesis areas?
Please tick one answer for each activity

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Original systematic reviews</th>
<th>Updating of systematic reviews</th>
<th>Clinical practice guidelines (CPGs)</th>
<th>Health technology assessments (HTAs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>11-20%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>21-30%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>31-40%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>41-50%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>51-60%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>61-70%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>71-80%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>81-90%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>91-100%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Not sure</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Not applicable</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

45. What is your organization/group's estimated budget generated/expended annually on the following research synthesis areas in US Dollars?
Please tick one option for each activity

<table>
<thead>
<tr>
<th>Budget Range</th>
<th>Original systematic reviews</th>
<th>Updating of systematic reviews</th>
<th>Clinical practice guidelines (CPGs)</th>
<th>Health technology assessments (HTAs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$11,000-40,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$41,000-70,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$71,000-$100,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$100,000-$200,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$200,000-$500,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>$500,000-$1,000,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>&gt;$1,000,000</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Not sure</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Not applicable</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

46. In what capacity have you previously participated in a systematic review?
Please tick all that apply

☐ Lead author of a systematic review
☐ Co-author of a systematic review
☐ Information specialist
☐ Statistician
☐ Methodologist/Epidemiologist
☐ Clinical expert
☐ Editor
☐ Project Manager/Coordinator
☐ None
☐ Other research capacity (please specify)
47. Please describe your training in systematic reviews.
   Please tick all that apply

   - University-level training or research in systematic reviews
   - Continuing education course(s) in systematic reviews
   - Workshop(s) in systematic reviews
   - Lecture(s) in systematic reviews
   - No training in systematic reviews
   - Other (Please specify)

48. How often do you utilize evidence from systematic reviews in the following capacities?
   Please tick one of the following options

<table>
<thead>
<tr>
<th>Daily</th>
<th>Weekly</th>
<th>Monthly</th>
<th>Never</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>To inform clinical practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To inform policy-making</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To inform funding decisions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For research purposes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

49. Would you like to receive a summary of this research when it is available?
   Please tick one of the following options

   - Yes
   - No

50. May we contact your organization/group within the next 12 months for more in depth information on its updating experiences? Please tick one of the following options

   - Yes
   - No

51. Please enter your contact information below. This will remain confidential and will only be viewed by the primary researcher.

   Responses are required for each field below:
   First & Last Name:
   Email address:
   Name of Organization/Group:
   Title or position:
   Degrees held:
   Numbers of Yrs with the Organization/Group:
   Street Address of the Organization/Group:
   City:
   Province/State/Region:
   Country:
   Postal/Zip Code:
   Website address of your organization:

   You have now completed the survey! As a token of our appreciation for your time, an Amazon voucher will be sent to you shortly.

   On behalf of the University of Ottawa EPC, we thank you!