Draft Methods Research Report

Number xx

Surveillance and Identification of Triggers for Updating Systematic Review: Implementation and Early Experience

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.:

Prepared by:

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by EPC I and EPC II under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, 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 health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for 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.

Public Domain Notice

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Disclaimer Regarding 508-Compliance

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [insert program e-mail address].

Financial Disclosure Statement

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: <Authors>. <Topic>. Evidence Report/Technology Assessment. No. <#>. (Prepared by <EPC Name> under Contract No. <##>.) Rockville, MD: Agency for Healthcare Research and Quality. <Month, Year>. http://www.ahrq.gov/clinic/epcix.htm. Accessed: <Date>. <URL>.

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 and strategies. 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 improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

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. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. 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.hhs.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

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Suchitra Iyer, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Peer Reviewers

[Spaceholder]

Surveillance and Identification of Triggers for Updating Systematic Reviews

Structured Abstract

Background: The question of how to determine when a systematic review needs to be updated is of considerable importance. Changes in the evidence can have significant implications for clinical practice guidelines and for clinical and consumer decision-making that depend on up-to-date systematic reviews as their foundation. Methods have been developed for assessing signals of need for updating, but these have been applied in research studies. To date there has been no application to a program of systematic reviews.

Objectives: The Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice (EPC) program commissioned two EPCs with additional assistance from another EPC with 42 topics to develop and implement an efficient surveillance system for identifying Comparative Effectiveness Reviews (CERs) in need for updating.

Approach: The Surveillance Program established a set of methods in assessing the needs for updating for the 14 CERs which was designed to be implemented in six month cycles. If a CER was deemed to be not up-to-date as it is a "high" priority for updating, it would fall off the cycle. The methods were based on methods developed by the two EPCs for prior projects on identifying signals for updating and included the following components: Identifying New Evidence from Published Studies, Identifying New Evidence from FDA (US), Health Canada, and Medicines and Healthcare products Regulatory Agency (MHRA) (UK), Identifying New Evidence from Experts and Expert Opinion, Checking for qualitative and quantitative signals. Once abstraction of the study conditions and findings for each new included study was completed and expert opinions were received, an assessment on whether the new findings provided a signal according to Method I and/or Method II, on a conclusion-by-conclusion basis. Lastly, an overall summary assessment was made that classified each CER as being of "high", "medium", or "low" priority for updating.

Between June 2011-June 2012 we began and completed the evaluation of 14 CERs. It was possible to complete all 14. Of the 14 CERs, 2 were classified as "high" priority, 3 as "medium" priority, and nine as "low" priority. Of the six CERs released prior to 2010 (meaning greater than 18 months prior to start of the Surveillance Program) two were judged as being "high" priority, two were judged as being "medium" priority, and two were judged as being "low" priority for updating. Of the remaining eight only one was judged as being "medium" priority for updating. All but one of the CERs released within the year prior to the start of the Surveillance Program were judged as being "low" priority.

Conclusions: We have established a Surveillance Program that evaluated 14 CERs over the course of 12 months. We have shown it is feasible to do such surveillance, in real-time, across a program that produces a large number of systematic reviews on diverse topics.

Contents

Executive Summary ES-1
Introduction ES-1
Methods ES-2
Results ES-4
Discussion ES-6
Conclusions ES-7
References ES-7
Introduction 1
Background 1
Methods 3
Identifying New Evidence from Published Studies 3
Identifying New Evidence from FDA (US), Health Canada, and Medicines and Healthcare products Regulatory Agency (MHRA) (UK) 3
Identifying New Evidence from Experts and Expert Opinion 4
Check for qualitative and quantitative signals 7
Determining Priority for Updating a CER 8
Summary Dissemination of Reports to AHRQ 9
Peer Review and Public Commentary 9
Results 10
Time to Complete the Surveillance Reports 10
Assessment Findings 14
Discussion 18
Conclusions 19
References 20
Abbreviations 23

Figures

Figure A. CER Surveillance Program ES-1
Figure 1. CER Surveillance Program 2
Figure 2. Sample Questionnaire Matrix 5
Figure 3. Method I 8

Tables

Table A. Method I and II ES-3
Table B. Fourteen Completed CER Topics ES-5
Table 1. Method I and II 7
Table 2. Fourteen Completed CER Topics 11
Table 3. 28 CER Topics Not Completed 12
Table 4. The descriptive statistics of the assessment findings for 14 CERs 14
Table 5. Characteristics of 14 comparative effectiveness reviews and their updating surveillance assessments 15

Appendixes

Appendix A. Examples of “Low”, “Medium”, and “High” Assessments
Appendix B. Methods for Identifying Regulatory Information or Safety Alerts

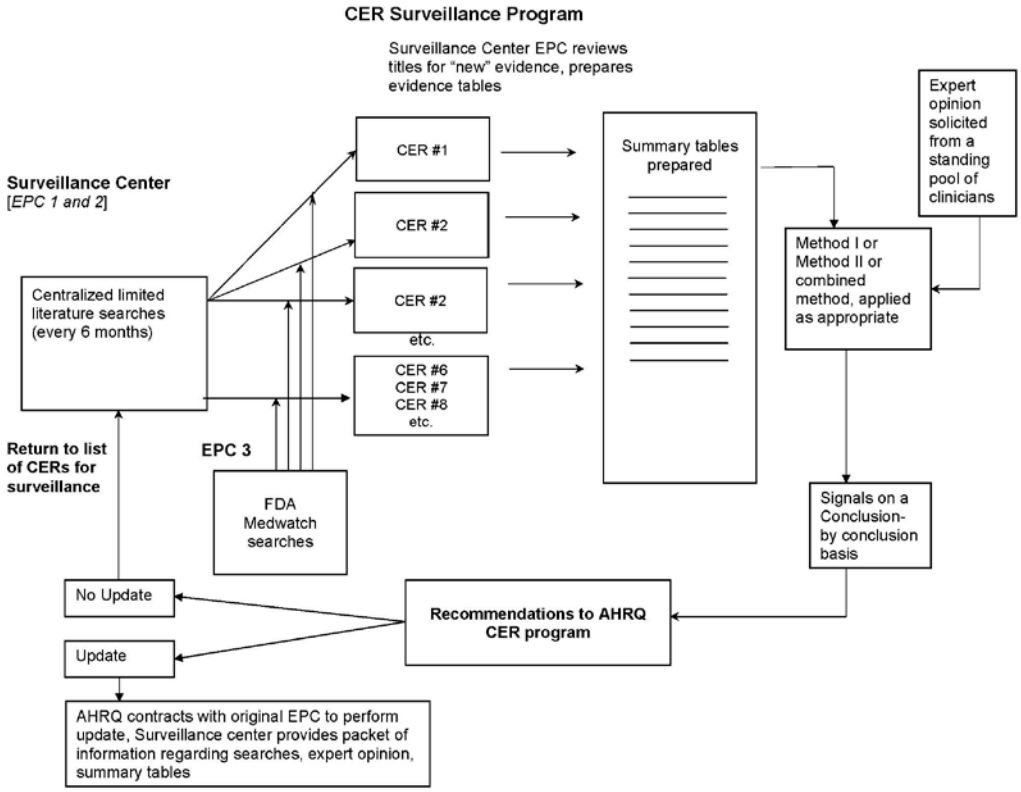
Executive Summary

Introduction

The Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice (EPC) program commissioned two EPCs with additional assistance from another EPC with 42 topics to develop and implement an efficient surveillance system for identifying Comparative Effectiveness Reviews (CERs) in need for updating. The two EPCs successfully developed a Surveillance Program to assess AHRQ's published CERs six months after their release date. The report describes the processes and methods that were used to develop the Surveillance Program using a sample of 14 completed assessments covering the period from June 2011-June 2012.

Figure A illustrates the overall process of the Surveillance Program developed and conducted by the two EPCs. Of the 42 CERs, EPC 1 was assigned 28 and EPC 2 was assigned 14..

Figure A. CER Surveillance Program



Methods

The Surveillance Program, using prior experience from the 2009 assessment project,[1] established a set of methods in assessing the needs for updating for the 14 CERs which was designed to be implemented in six month cycles. If a CER was deemed to be not up-to-date as it is a "high" priority for updating, it would fall off the cycle. The methods that were used are described below:

Identifying New Evidence from Published Studies

A limited search and study selection of those searches were used in finding new studies. Across all topics, a literature search strategy of the original CER six to twelve months prior to the end date of the original CER search in order to ensure overlap between the searches which included searching Medline/Pubmed and Cochrane, and on a topic specific basis additional databases was conducted. The search included the five general medical interest journals (Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Lancet, and the New England Journal of Medicine) and the most relevant specialty journals specific to that topic. The specialty journals were those most highly represented among the references for the original report.

For the study selection, the same inclusion and exclusion criteria were used from the original CER. A single reviewer, experienced in systematic reviews, conducted a screening of the titles and abstracts and requested any articles deemed relevant to the topic. From those articles, a single reviewer extracted relevant data from accepted articles.

Identifying New Evidence from FDA (US), Health Canada, and Medicines and Healthcare products Regulatory Agency (MHRA) (UK)

An additional EPC, under contract from AHRQ, monitored at monthly intervals the U.S.Food and Drug Administration (FDA), Health Canada, and MHRA websites for any new regulatory information or safety alerts about drugs relevant to the included CERs. This information was forwarded to the two EPCs as appropriate. Appendix B outlines the methods used by the EPC to identify new regulatory information or safety alerts.

Identifying New Evidence from Experts and Expert Opinion

For each of the CERs that were assessed, a questionnaire matrix that listed the key conclusions from the original executive summary was created. Experts were asked whether each listed conclusion was still valid, and if not, to provide information about new evidence. The questionnaire matrix was sent to experts in the field, including the original project leader, technical expert panel (TEP) members, and peer reviewers.

Check for qualitative and quantitative signals

Once abstraction of the study conditions and findings for each new included study was completed and expert opinions were received, an assessment on whether the new findings provided a signal according to Method I and/or Method II, on a conclusion-by-conclusion basis. Table A lists the criteria used for reaching one and/or the other conclusions.[1, 2]

Table A. Method I and II

	Method I
	Qualitative Criteria for Potentially Invalidating Signals
A1	Opposing findings: A pivotal* trial or systematic review (or guidelines) including at least one new trial that characterized the treatment in terms opposite to those used earlier.
A2	Substantial harm: A pivotal trial or systematic review (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making.
A3	A superior new treatment: A pivotal trial or systematic review (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.
	Qualitative Criteria for Signals of Major Changes
A4	Important changes in effectiveness short of "opposing findings"
A5	Clinically important expansion of treatment
A6	Clinically important caveat
A7	Opposing findings from discordant meta-analysis or nonpivotal trial
	Quantitative Criteria Signals of Changes in Evidence
B1	A change in statistical significance (from nonsignificant to significant)
B2	A change in relative effect size of at least 50 percent
	Method II Indications for the Need for an Update
1	Original conclusion is still valid and this portion of the original report does not need Updating. This conclusion was reached if we found no new evidence or only confirmatory evidence and all responding experts assessed the CER conclusion as still valid, we classified the CER conclusion as still valid.
2	Original conclusion is possibly out of date and this portion of the original report may need updating. This conclusion was reached if we found some new evidence that might change the CER conclusion, and /or a minority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as possibly out of date.
3	Original conclusion is probably out of date and this portion of the original report may need updating. This conclusion was reached if we found substantial new evidence that might change the CER conclusion, and/or a majority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as probably out of date.
4	Original conclusion is out of date. This conclusion was reached if we found new evidence that rendered the CER conclusion out of date or no longer applicable, we classified the CER conclusion as out of date. Recognizing that our literature searches were limited, we reserved this category only for situations where a limited search would produce prima facie evidence that a conclusion was out of date, such as the withdrawal of a drug or surgical device from the market, a black box warning from FDA, etc.

*A pivotal trial was defined as: a trial published in top 5 general medical journals (Annals of Intern Med, BMJ, JAMA, The Lancet, and NEJM), or a trial not published in the above top 5 journals but had a sample size of at least triple the size of the previous largest trial in the original CER.

Determining Priority for Updating a CER

The above classifications were made on a conclusion-by-conclusion basis. We next needed to determine an overall judgment of the priority for updating based on the degree to which the current CER's conclusions were out-of-date. We used two criteria in making our final conclusion for a CER:

- How much of the CER is possibly, probably, or certainly out of date?
- How out of date is that portion of the CER? For example, would the potential changes to the conclusions involve refinement of original estimates or do the potential changes mean some therapies are no longer favored or may not exist? Is the portion of the CER that is probably or certainly out of date an issue of safety (a drug withdrawn from the market, a black box warning) or the availability of a new drug within class (the latter being less of a signal to update than the former)?

This final conclusion was a global judgment made by all the individuals working on each particular CER. We classified CERs as being "low", "medium", or "high" priority for updating, with a notation explaining the rationale for "high" priority updates.

Results

Between June 2011-June 2012 we began and completed the evaluation of 14 CERs. Of the 14, CERs 2 were classified as "high" priority, 3 as "medium" priority, and nine as "low" priority. Table B shows the results of those 14 CERs assessed and presents the number of days to complete each of the 14 surveillance reports. The mean was 86 days and the median was 74 days, with the majority of reports being completed between 65 and 102 days. The assessment of 14 CERs that were submitted by the two EPCs can be found on AHRQ's website: www.ahrq.gov.

Of the six CERs released prior to 2010 (meaning greater than 18 months prior to start of the Surveillance Program) two were judged as being "high" priority, two were judged as being "medium" priority, and two were judged as being "low" priority for updating. Of the remaining eight only one was judged as being "medium" priority for updating. All but one of the CERs released within the year prior to the start of the Surveillance Program were judged as being "low" priority.

Table B. Fourteen Completed CER Topics

CER Topic	Release date	Date sent to AHRQ	Days to complete	Priority
Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer (CER 13)	2/1/2008	5/18/2012	74	**High**
Comparative Effectiveness of Radiofrequency Catheter Ablation for Atrial Fibrillation (CER 15)	7/6/2009	11/30/11	69	Low
Comparative Effectiveness of Lipid-Modifying Agents (CER 16)	9/1/2009	12/23/11	72	**High**
Comparative Effectiveness of Medications To Reduce Risk of Primary Breast Cancer in Women (CER 17)	9/14/2009	11/08/11	119	Medium
Comparative Effectiveness of Angiotensin Converting Enzyme Inhibitors or Angiotensin II Receptor Blockers Added to Standard Medical Therapy for Treating Stable Ischemic Heart Disease (CER 18)	10/16/2009	12/23/11	38	Low
Comparative Effectiveness of Core Needle Biopsy and Open Surgical Biopsy for Diagnosis of Breast Lesions (CER 19)	12/15/2009	12/16/11	102	Medium
Comparative Effectiveness and Safety of Radiotherapy Treatments for Head and Neck Cancer (CER 20)	5/27/2010	11/4/11	75	Medium
Comparative Effectiveness of In-Hospital Use of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care (CER 21)	6/1/2010	2/24/12	65	Low
Comparative Effectiveness of Nonoperative and Operative Treatments for Rotator Cuff Tears (CER 22)	6/5/2010	2/24/12	65	Low
Effectiveness of Recombinant Human Growth Hormone (rhGH) in the Treatment of Patients With Cystic Fibrosis (CER 23)	10/4/2010	12/16/11	100	Low
Traumatic Brain Injury and Depression (CER 25)	4/13/2011	4/10/12	161	Low
Comparative Effectiveness of Therapies for Children with Autism Spectrum Disorders (CER 26)	4/4/2011	01/23/12	97	Low
Comparative Effectiveness of Pain Management Interventions for Hip Fracture (CER 30)	5/17/2011	4/10/12	133	Low
Comparative Effectiveness of Terbutaline Pump for the Prevention of Preterm Birth (CER 35)	9/29/2011	5/4/12	38	Low

Discussion

Our preliminary results indicate that a small proportion of CERs may be in need of urgent updating 1-3 years after their last search date. Of the 14 CERs assessed between June 2011 and June 2012, 9 (64%) were classified as having low priority for updating and 3 CERs (22%) had medium priority for updating. Two CERs (14%) were determined to have high priority for updating.

The implementation of the surveillance assessment program to determine the updating status of published AHRQ CERs has not been without challenges. The assessment of currency and validity of conclusions for each KQ of a CER was based on the totality of information compiled through multiple different sources including the qualitative/quantitative signals, expert opinion, and FDA alerts. Although we used operational and standardized definitions throughout the process to ensure relative consistency in the assessments, human judgment is required to interpret the newly identified evidence in relation to the conclusion of the CER. This judgment is a potential source of variability in the assessment. However, our prior work has shown that in at least one explicit assessment, the inter-rater reliability of these judgments was at least moderate.[3]

An additional challenge was the variability in presentation among the original CERs. Not all CERs presented their executive summaries along with KQs and corresponding conclusions in an identical or even similar manner (e.g., degree of detail, format, or level of summarization). For example, in some CERs, conclusions for each KQ were stratified by the outcome and/or intervention, resulting in multiple conclusions. In other instances, executive summaries were not sufficiently detailed to be able to extract a specific clearly formulated conclusion, and therefore, the reviewers had to probe the entire body of text in the CER. Moreover, some conclusions in KQs were not readily amenable to updating with respect to comparative effectiveness. For example, these conclusions included descriptive information on prevalence of certain risk factors in specific populations.

There were a few instances when experts differed in their opinion or the experts' opinion differed from what was demonstrated in newly identified studies. Such observations may reflect the possibility that experts differ in knowledge of their respective content areas (AHRQ CERs enlist the input from a diverse set of technical experts for exactly this reason) as well as how up-to-date they are in terms of the emerging new literature.[4] Additionally, for most CERs we received back responses from fewer than 50 percent of the experts originally contacted, although since the experts were not a random sample the implications of this low response rate are not as clear.

To our knowledge, this surveillance assessment project is the first large scale effort that has applied methods to assess the updating status of evidence-based reports (or systematic reviews) in a structured and standardized manner. The application of these methods has proven to be relatively feasible, efficient and at the same time a comprehensive and systematic approach for assessing the need for updating individual CERs across a wide range of health interventions.

However, it is premature to generalize our findings to a broader population of CERs or systematic reviews, and more data collection is necessary.

Conclusions

We have established a Surveillance Program that evaluated 14 CERs over the course of 12 months. Regarding the need to update, two were classified as “high” priority, three as “medium” priority, and nine a “low: priority, 9 to 55 months after the last search date of the original CER. We have shown that a program for regular and active surveillance of Comparative Effectiveness Reviews is feasible.

References

1. Shekelle PG, Newberry SJ, Maglione M, et al. Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005-2009) (Prepared by the Southern California Evidence-based Practice Center). Rockville, MD: Agency for Healthcare Research and Quality; October 2009.

2. Shojania KG, Sampson M, Ansari MT, et al. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007 Aug 21;147(4):224-33. PMID 17638714.

3. Chung M, Newberry SJ, Ansari MT, et al. Two methods provide similar signals for the need to update systematic reviews. J Clin Epidemiol. 2012 Jun;65(6):660-8. PMID 22464414.

4. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. Ann Intern Med. 2011;154(1):50-5.

Introduction

Background

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

Systematic reviews are commonly updated at a pre-set time after publication.[2] For example, since 2002, the Cochrane Collaboration's policy has been to update Cochrane reviews every 2 years.[3] Such updates involve an investment of time and effort that may not be appropriate for all topics. In 2005, 254 Cochrane updates performed in 2002 were compared with reviews from 1998. Only 23 (9%) had a change in conclusion, which supports use of a priority approach, rather than an automatic time-based approach, to determine the need for an update.[4]

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

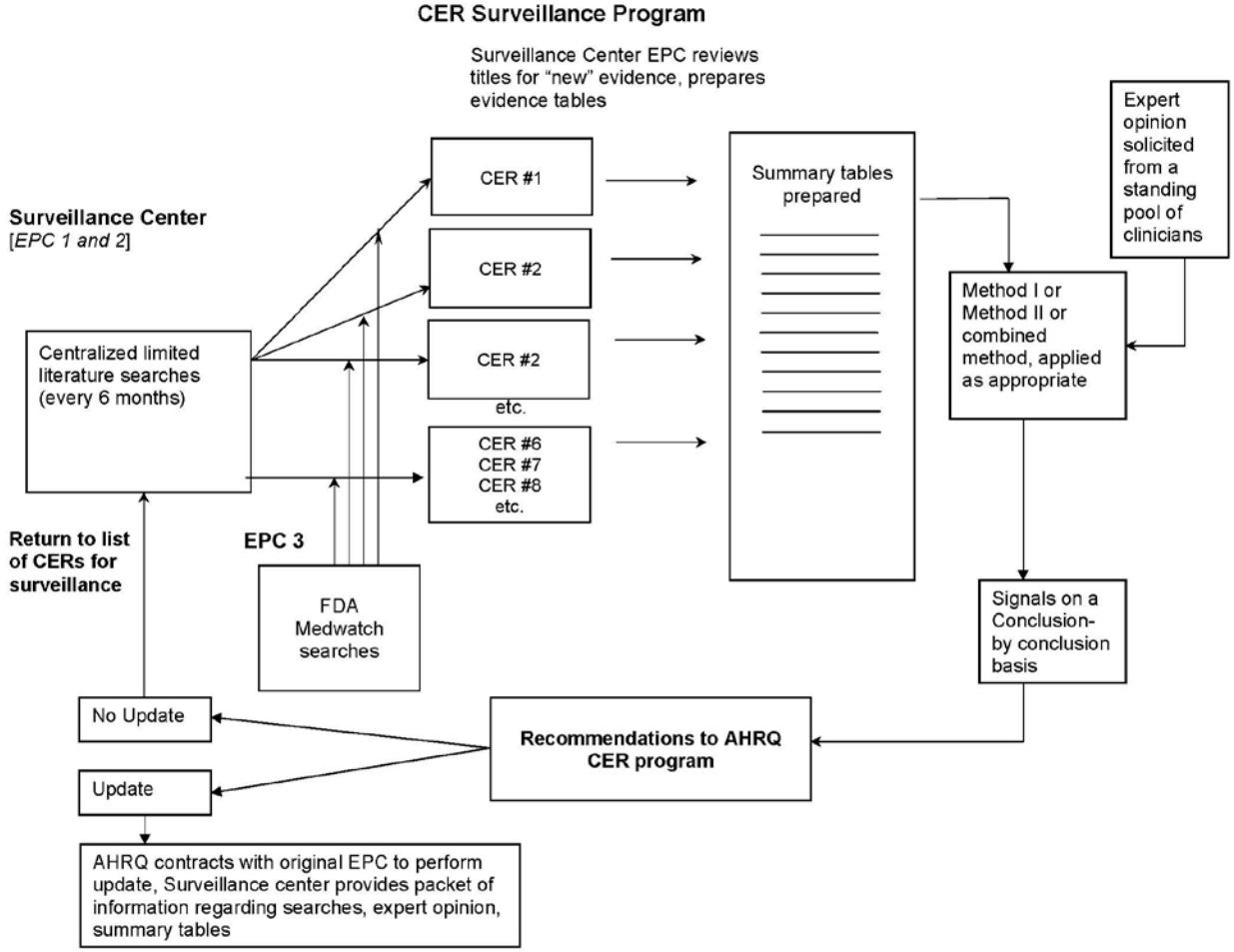
A series of subsequent methods projects led to the development of the Surveillance Program. In early 2008, AHRQ determined that to meet their intended objectives the EHC Program should assess the need for the CERs completed to that point to be updated. The EPC was tasked with conducting this assessment. As part of this project, the EPC proposed a model for a program of regular surveillance for AHRQ Comparative Effectiveness Reviews.[7]

In 2010, AHRQ commissioned a pilot study to compare the results of Method I and II for identifying signals for the need for updating. Chosen as test cases were three evidence reports on

omega-3 fatty acids (omega-3 FA): the effects of omega-3 FA for preventing and treating neurological disorders[10]; the effects of omega-3 FA for preventing and treating cancer[11]; and the effects of omega-3 FA on risk factors and intermediate markers for cardiovascular disease (CVD).[12] The report concluded that the data support the use of either method, as in general they provide similar signals for the possible need to update systematic reviews.[13, 14] Additionally, the report hypothesized that a hybrid model may offer advantages over either individual model.

AHRQ then commissioned the current Surveillance Program to evaluate 42 CERs using Method I and/or Method II for identifying signals indicating the need for updating.
Figure 1 illustrates the overall process of the Surveillance Program developed and conducted by the two EPCs. Of the 42 CERs, EPC 1 was assigned 28 and EPC 2 was assigned 14.

Figure 1. CER Surveillance Program



Methods

AHRQ commissioned two EPCs with additional assistance from another EPC with 42 topics to develop and implement an efficient surveillance system for identifying CERs in need for updating. This report covers the period from June 2011-June 2012 and summarizes the assessment of the 14 CERs.

Identifying New Evidence from Published Studies

Search Strategy

The surveillance assessment system was designed to be implemented in 6 month cycles. The first cycle of this process starts with the assessment of a CER 6 months after its publication on the AHRQ website. The CERs determined to be up to date in the first cycle are re-assessed 6 months later in the next cycle. The CERs determined to be clearly out of date in the first cycle are not re-assessed (no next cycle).

Starting with the search strategy employed for the original report, we conducted a limited literature search that included at least Medline/Pubmed and Cochrane, and on a topic specific basis additional databases. The search included the five general medical interest journals (Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Lancet, and the New England Journal of Medicine) and the most relevant specialty journals specific to that topic. The specialty journals were those most highly represented among the references for the original report. In general, we followed the search strategy from the original CER. However, we did make some modifications. For example, if we knew of new drugs for the condition, these were added to the search terms. Searches were started six to twelve months prior to the end date of the original CER search in order to ensure overlap between the searches.

Study Selection

In general, we used the same inclusion and exclusion criteria as the original CER. A single reviewer, experienced in systematic reviews, conducted a screening of the titles and abstracts and requested any articles deemed relevant to the topic. From those articles, a single reviewer extracted relevant data from accepted articles.

Identifying New Evidence from FDA (US), Health Canada, and Medicines and Healthcare products Regulatory Agency (MHRA) (UK)

An additional EPC, under contract from AHRQ, monitored at monthly intervals the U.S.Food and Drug Administration (FDA), Health Canada, and MHRA websites for any new regulatory information or safety alerts about drugs relevant to the included CERs. This information was forwarded to the two EPCs as appropriate.

Identifying New Evidence from Experts and Expert Opinion

Across all topics, a questionnaire matrix that listed the key conclusions from the original executive summary was created. Experts were asked whether each listed conclusion was still valid, and if not, to provide information about new evidence (see Figure 2 below). The questionnaire matrix was sent to experts in the field, including the original project leader, technical expert panel (TEP) members, and peer reviewers.

Figure 2. Sample Questionnaire Matrix

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Key Question 1: What is the prevalence of depression after traumatic brain injury, and does the area of the brain injured, the severity of the injury, the mechanism or context of injury, or time to recognition of the traumatic brain injury or other patient factors influence the probability of developing clinical depression?			
The prevalence of depression after traumatic brain injury is approximately 30 percent across multiple time points up to and beyond a year. Based on structured clinical interviews, on average 27 percent of TBI patients met criteria for depression 3 to 6 months from injury; 32 percent at 6 to 12 months; and 33 percent beyond 12 months. Higher prevalence is reported in many study populations. No strong predictors are available to select a screening window or to advise TBI patients or their providers about risk of depression.	☐	New Evidence:	☐
Key Question 2: When should patients who suffer traumatic brain injury be screened for depression, with what tools, and in what setting?			
Prevalence of depression is high at multiple time points after TBI. No evidence provides a basis for targeting screening to one timeframe over another. Likewise, the literature is insufficient to determine whether tools for detecting depression that have been validated in other populations can accurately identify depression in individuals with TBIs. Nor does the literature support any one tool over the others.	☐	New Evidence:	☐
Key Question 3: Among individuals with TBI and depression, what is the prevalence of concomitant psychiatric/behavioral conditions, including anxiety disorders, post-traumatic stress disorder (PTSD), substance abuse, and major psychiatric disorders?			
When conditions were reported individually, anxiety disorder was most prevalent and affected from 31 to 61 percent of study	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
participants in four papers. PTSD, a major anxiety disorder, was observed in 37 percent of depressed patients and in no patients without depression, and panic disorder was seen in 15 percent of patients with major depression, but not measured in those without depression. Consideration of potential for coexisting psychiatric conditions is warranted.			
Key Question 4: What are the outcomes (short and long term, including harm) of treatment for depression among traumatic brain injury patients utilizing psychotropic medications, individual/group psychotherapy, neuropsychological rehabilitation, community-based rehabilitation, complementary and alternative medicine, neuromodulation therapies, and other therapies?			
Only two publications addressed treatment for individuals diagnosed with depression after a traumatic brain injury: Both were studies of antidepressant efficacy (one a controlled trial of sertraline and one an open-label trial of citalopram). The sertraline trial showed no significant effect compared with placebo, and the citalopram study did not show improvement in a majority of participants.	☐	New Evidence:	☐
Key Question 5: Where head-to-head comparisons are available, which treatment modalities are equivalent or superior with respect to benefits, short- and long-term risks, quality of life, or costs of care?			
No head-to-head trials were identified that compared the effectiveness of two or more modalities for treating depression that follows TBI. Such studies are needed.	☐	New Evidence:	☐
Key Question 6: Are the short- and long-term outcomes of treatment for depression after TBI modified by individual characteristics, such as age, preexisting mental health status or medical conditions, functional status, and social support?			
No studies were identified that assessed the impact of demographic or other potentially modifying characteristics on treatment effectiveness. Future research needs to address this issue.	☐	New Evidence:	☐
Are there new data that could inform the key questions that might not be addressed in the conclusions?			

Check for qualitative and quantitative signals

Once abstraction of the study conditions and findings for each new included study was completed and expert opinions were received, we assessed whether the new findings provided a signal according to Method I and/or Method II, on a conclusion-by-conclusion basis. Table 1 lists the criteria used for reaching conclusions.[7, 9]

Table 1. Method I and II

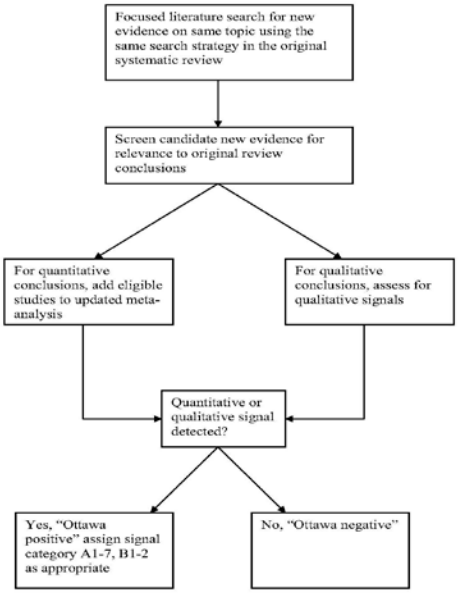
	Method I
	Qualitative Criteria for Potentially Invalidating Signals
A1	Opposing findings: A pivotal* trial or systematic review (or guidelines) including at least one new trial that characterized the treatment in terms opposite to those used earlier.
A2	Substantial harm: A pivotal trial or systematic review (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making.
A3	A superior new treatment: A pivotal trial or systematic review (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.
	Qualitative Criteria for Signals of Major Changes
A4	Important changes in effectiveness short of "opposing findings"
A5	Clinically important expansion of treatment
A6	Clinically important caveat
A7	Opposing findings from discordant meta-analysis or nonpivotal trial
	Quantitative Criteria Signals of Changes in Evidence
B1	A change in statistical significance (from nonsignificant to significant)
B2	A change in relative effect size of at least 50 percent
	Method II Indications for the Need for an Update
1	Original conclusion is still valid and this portion of the original report does not need Updating. This conclusion was reached if we found no new evidence or only confirmatory evidence and all responding experts assessed the CER conclusion as still valid, we classified the CER conclusion as still valid.
2	Original conclusion is possibly out of date and this portion of the original report may need updating. This conclusion was reached if we found some new evidence that might change the CER conclusion, and /or a minority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as possibly out of date.
3	Original conclusion is probably out of date and this portion of the original report may need updating. This conclusion was reached if we found substantial new evidence that might change the CER conclusion, and/or a majority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as probably out of date.
4	Original conclusion is out of date. This conclusion was reached if we found new evidence that rendered the CER conclusion out of date or no longer applicable, we classified the CER conclusion as out of date. Recognizing that our literature searches were limited, we reserved this category only for situations where a limited search would produce prima facie evidence that a conclusion was out of date, such as the withdrawal of a drug or surgical device from the market, a black box warning from FDA, etc.

*A pivotal trial was defined as: a trial published in top 5 general medical journals (Annals of Intern Med, BMJ, JAMA, The Lancet, and NEJM), or a trial not published in the above top 5 journals but had a sample size of at least triple the size of the previous largest trial in the original CER.

Method I involved detection of qualitative and/or quantitative signals indicating the need for updating through the assessment of new evidence using specific categories for qualitative (A1-A7) and quantitative (B1-B2) signals as reported in Table 1. For example, a qualitative signal was considered a finding from a newly published pivotal trial which was opposite to that of the CER with respect to an efficacy outcome (e.g., effective vs. ineffective or vice-versa) or harm (e.g., risk of harm outweighs the previously observed benefits), a superior new treatment (e.g., new treatment significantly more effective than one assessed in the CER), or a new subgroup of population (the treatment assessed in the CER has been expanded to a new subgroup of participants). An example of a quantitative signal is the incorporation of a new trial (or trials) into a meta-analysis of the CER leading to an overturn of a statistically non-significant pooled estimate into a statistically significant one or vice-versa.

The specific steps involved in Method I are shown below in Figure 3.

Figure 3. Method I



We constructed a summary table that included the following for each key question: original conclusion, findings of the new literature search, summary of expert assessment, and findings from another EPCs search of regulatory bodies, and our final classification of the conclusion.

Determining Priority for Updating a CER

The above classifications were made on a conclusion-by-conclusion basis. We next needed to determine an overall judgment of the priority for updating based on the degree to which the current CER's conclusions were out-of-date. We used two criteria in making our final conclusion for a CER:

- How much of the CER is possibly, probably, or certainly out of date?

- How out of date is that portion of the CER? For example, would the potential changes to the conclusions involve refinement of original estimates or do the potential changes mean some therapies are no longer favored or may not exist? Is the portion of the CER that is probably or certainly out of date an issue of safety (a drug withdrawn from the market, a black box warning) or the availability of a new drug within class (the latter being less of a signal to update than the former)?

This final conclusion was a global judgment made by all the individuals working on each particular CER. We classified CERs as being "low", "medium", or "high" priority for updating, with a notation explaining the rationale for "high" priority updates.

Summary Dissemination of Reports to AHRQ

We developed a format for a short summary report that presents the findings from the surveillance process to AHRQ. This format includes a title page that lists the final classification ("low," "medium," or "high") of the priority for updating the CER, the details of the literature search and its yield (with evidence tables), the findings from FDA, Health Canada, and MHRA, the results of any expert opinion that was sought, and a summary table that contains each conclusion from the original CER and our assessment of the degree to which it maybe out-of-date. Examples of such reports (one each judged as being at "low," "medium," and "high" priority for updating) are included in Appendix A.

Peer Review and Public Commentary

Experts in [X, Y, and Z] fields and individuals representing stakeholder and user communities were invited to provide external peer review of this CER; AHRQ and an associate editor also provided comments. The draft report will be posted on the AHRQ website for 4 weeks to elicit public comment. We will address all reviewer comments, revising the text as appropriate, and will document everything in a "disposition of comments report" that will be made available 3 months after the Agency posts the final CER on the AHRQ website.

Results

Time to Complete the Surveillance Reports

Between June 2011-June 2012 we began and evaluated the evaluation of 14 CERs. Table 2 indicates the CERs that were evaluated and the priority they received, the date of their release, and the date of when either EPC 1 or EPC II sent the assessments to AHRQ. In addition, Table 2 presents the number of days to complete each of the 14 surveillance reports. The mean was 86 days and the median was 74 days, with the majority of reports being completed between 65 and 102 days.

Table 3 shows the remaining 28 CERs; four are in progress; eleven have not reached their assessment date; and thirteen have not yet been released.

The assessment of 14 CERs that were submitted by EPC I or EPC II can be found on AHRQ's website: www.ahrq.gov.

Table 2. Fourteen Completed CER Topics

CER Topic	Release date	Date sent to AHRQ	Days to complete	Priority
Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer (CER 13)	2/1/2008	5/18/2012	74	**High**
Comparative Effectiveness of Radiofrequency Catheter Ablation for Atrial Fibrillation (CER 15)	7/6/2009	11/30/11	69	Low
Comparative Effectiveness of Lipid-Modifying Agents (CER 16)	9/1/2009	12/23/11	72	**High**
Comparative Effectiveness of Medications To Reduce Risk of Primary Breast Cancer in Women (CER 17)	9/14/2009	11/08/11	119	Medium
Comparative Effectiveness of Angiotensin Converting Enzyme Inhibitors or Angiotensin II Receptor Blockers Added to Standard Medical Therapy for Treating Stable Ischemic Heart Disease (CER 18)	10/16/2009	12/23/11	38	Low
Comparative Effectiveness of Core Needle Biopsy and Open Surgical Biopsy for Diagnosis of Breast Lesions (CER 19)	12/15/2009	12/16/11	102	Medium
Comparative Effectiveness and Safety of Radiotherapy Treatments for Head and Neck Cancer (CER 20)	5/27/2010	11/4/11	75	Medium
Comparative Effectiveness of In-Hospital Use of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care (CER 21)	6/1/2010	2/24/12	65	Low
Comparative Effectiveness of Nonoperative and Operative Treatments for Rotator Cuff Tears (CER 22)	6/5/2010	2/24/12	65	Low
Effectiveness of Recombinant Human Growth Hormone (rhGH) in the Treatment of Patients With Cystic Fibrosis (CER 23)	10/4/2010	12/16/11	100	Low
Traumatic Brain Injury and Depression (CER 25)	4/13/2011	4/10/12	161	Low
Comparative Effectiveness of Therapies for Children with Autism Spectrum Disorders (CER 26)	4/4/2011	01/23/12	97	Low
Comparative Effectiveness of Pain Management Interventions for Hip Fracture (CER 30)	5/17/2011	4/10/12	133	Low
Comparative Effectiveness of Terbutaline Pump for the Prevention of Preterm Birth (CER 35)	9/29/2011	5/4/12	38	Low

Table 3. 28 CER Topics Not Completed

CER Topic	Release Status	Status of EPC I or EPC II
Effectiveness of Disease-modifying Anti-rheumatic drugs (DMARDs) in Children with Juvenile Idiopathic Arthritis (JIA) (CER 28)	9/26/2011	Assessment is in progress by EPC I
Effectiveness of Early Diagnosis, Prevention, and Treatment of Clostridium difficile-Associated Disease (CER 31)	12/19/2011	Expected start date of assessment by the EPC I: June 2012
Comparative Effectiveness Review of the Diagnosis and Treatment of Obstructive Sleep Apnea in Adults (CER 32)	8/5/2011	Assessment is in progress by EPC I
Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults (CER 33)	9/23/2011	Assessment is in progress by EPC I
Diagnosis and Comparative Effectiveness of Treatments for Urinary Incontinence in Adult Women (CER 36)	4/9/2012	Expected start date of assessment by EPC II: October 2012
Screening for and Management of Chronic Kidney Disease Stages 1-3 (CER 37)	1/31/2012	Expected start date of assessment EPC I: July 2012
Comparative Effectiveness of First and Second Generation Antipsychotics in the Pediatric and Young Adult Populations (CER 39)	2/20/2012	Expected start date of assessment by EPC I: August 2012
Comparative Effectiveness of Treatments for Chronic Pelvic Pain in Women (CER 41)	1/31/2012	Expected start date of assessment by EPC I: July 2012
Comparative Effectiveness of Adjunctive Devices to Remove Thrombus and/or Protect from Distal Embolization in ACS Patients Undergoing PCI of Native Vessels (CER 42)	12/29/2011	Expected start date of assessment by EPC I: June 2012
Attention Deficit Hyperactivity Disorder: Effectiveness of Treatment in At-risk Preschoolers; Long-term Effectiveness in all Ages; and Variability in Prevalence, Diagnosis, and Treatment (CER 44)	10/21/2011	Assessment is in progress by EPC I
Comparative Effectiveness of Self-measured Blood Pressure Monitoring (CER 45)	1/31/2012	Expected start date of assessment by EPC II: July 2012
Comparative Effectiveness of Hematopoietic Stem-cell Transplantation in the Pediatric Population (CER 48)	2/1/2012	Expected start date of assessment by EPC II: August 2012
Comparative Effectiveness of Venous Thromboembolism (VTE) Prophylaxis in Orthopedic Surgery (CER 49)	3/13/2012	Expected start date of assessment by EPC II: September 2012
Effectiveness of Antinuclear Antibody, Rheumatoid Factor, and Cyclic-Citrullinated Peptide Testing for the Evaluation of Musculoskeletal Complaints in Pediatric Populations (CER 50)	3/7/2012	Expected start date of assessment by EPC I: September 2012
Dietary Supplements in the Elderly (CER 51)	4/30/2012	Expected start date of assessment by EPC I: October 2012
Pharmacologic Therapy vs Medical Foods for the Treatment of Phenylketonuria (PKU) (CER 56)	2/22/2012	Expected start date of assessment by EPC I: August 2012
Comparative Effectiveness of Screening for Glaucoma (CER 59)	4/12/2012	Expected start date of assessment by EPC I: October 2012
Comparative Effectiveness of Treatment for Open-Angle Glaucoma (CER 60)	4/12/2012	Expected start date of assessment by EPC I: October 2012

CER Topic	Release Status	Status of EPC I or EPC II
Comparative Effectiveness of Treatment for Depression after Unsatisfactory Response to SSRIs (CER 62)	4/26/2012	Expected start date of assessment by EPC I: October 2012
Progestogens for Prevention of Preterm Birth	Not released	Waiting for the report to be released
Comparative Effectiveness of First and Second Generation Antipsychotics in the Adult Population	Not released	Waiting for the report to be released
Comparative Effectiveness of Pharmacologic Therapies for the Management of Crohn's Disease	Not released	Waiting for the report to be released
Comparative Effectiveness of Allergen-specific Immunotherapy for the Treatment of Allergic Rhinoconjunctivitis and/or Asthma	Not released	Waiting for the report to be released
Comparative Effectiveness of Bariatric Surgery and Non-Surgical Therapy in Adults with Metabolic Conditions and Body Mass Index of 30 to 34.9	Not released	Waiting for the report to be released
Comparative Effectiveness of Breathing Exercises and/or Retraining Techniques in the Treatment of Asthma	Not released	Waiting for the report to be released
Comparative Effectiveness of Procalcitonin-Guided Decisions for the Diagnosis and Management of Sepsis	Not released	Waiting for the report to be released
Comparative Effectiveness and Safety of Multiple Daily Injections or Insulin Pump Therapy With or Without Continuous Glucose Monitoring for Diabetes Mellitus	Not released	Waiting for the report to be released
Comparative Effectiveness of Surgical Options for Inguinal Hernia	Not released	Waiting for the report to be released

Assessment Findings

Of the 14 CERs, four (29%) were up-to-date in the 9 to 55 months following their original search date (CER# 22, 23, 25, 35). For ten (71%) CERs (CER#s 13, 15, 16, 17, 18, 19, 20, 21, 26, 30), at least one conclusion within a KQ changed the status from 'up to date' to 'probably/possibly out of date' or 'out of date.' For four (29%) CERs (CER# 13, 18, 19, 20), all conclusions within at least one KQ changed the status from 'up to date' to 'probably/possibly out of date' or 'out of date.' The characteristics and descriptive statistics of the 14 CERs and the corresponding surveillance assessment results are presented in Table 4 and Table 5.

Of the 14 CERs, two (14.3%; CER #13 and 16), three (21.4%; CER #17, 19, 20), and nine (64.4%; CER # 15, 18, 21, 22, 23, 25, 26, 30, 35) were assigned to 'high,' 'medium,' and 'low' priority for updating, respectively. Of the six CERs released prior to 2010 (meaning greater than 18 months prior to start of the Surveillance Program) two were judged as being "high" priority, two were judged as being "medium" priority, and two were judged as being "low" priority for updating. Of the remaining eight only one was judged as being "medium" priority for updating. All but one of the CERs released within the year prior to the start of the Surveillance Program were judged as being "low" priority.

The experts' response rate ranged from 20 percent to 100 percent with a median of 34.5 percent.

None of the 14 CERs for which we performed the surveillance assessments had an FDA black box warning associated with an agent, device, or procedure that was a topic of the CER (the strongest FDA warning indicated a significant risk of a serious or even life-threatening adverse effect). Five CERs had safety communications, adverse effects, and label change alerts (CER #s, 16, 18, 23, 25, 26), none of which was sufficient to impact the updating priority of those CERs. CER#16 had a total of six label change alerts and 3 drug safety communication alerts, CER#18 had three alerts, and three CERs (#s 23, 25, and 26) had one alert each.

Table 4. The descriptive statistics of the assessment findings for 14 CERs

# of key questions in CER	Median (range) of studies in CER	Median # (range) of new potentially eligible studies	# of months between the last search date in CER and update search	Proportion of key questions with all conclusions up to date	Proportion of key questions with all conclusions possibly out of date
3 (CER#19, 16) -7 (CER#23,26)	107.5 (14- 436)	15 (0 -33)	9 (CER# 25)- 55 (CER#13)	25% (CER#s 20, 13) - 100% (CER#s 22,23,25,35)	14% (CER#18)- 33% (CER#19)

Table 5. Characteristics of 14 comparative effectiveness reviews and their updating surveillance assessments

CER Title (AHRQ #) Author name (publication date) [Journal publication, if available]	Latest search date for CER (across databases)	# of included studies in CER (total or per KQ)	Period covered by update search	# Of included studies in assessment report	Number of conclusions within the key questions in CER by updating status								Updating priority for the CER		
					KQ#	# Conclusions Up-to-date	KQ#	# Conclusions Probably out of date	KQ#	# Conclusions Possibly out of date	KQ#	# Conclusions Out of date	Low	Medium	High
Assessed by EPC 1															
Comparative Effectiveness of Core Needle Biopsy and Open Surgical Biopsy for Diagnosis of Breast Lesions (19) Bruening (December, 2009)[15] [Systematic review: comparative effectiveness of core-needle and open surgical biopsy to diagnose breast lesions][16]	September 2009	107 (KQ1-2) NA (KQ3)	March 2009 to September 2011	19	1 2	11/16 3/4			1a 2 3	4/16 1/4 all	1	1/16		x	
Comparative Effectiveness of Medications To Reduce Risk of Primary Breast Cancer in Women (17) Nelson (September 2009)[17] [Systematic review: comparative effectiveness of medications to reduce risk for primary breast cancer][18]	January 2009	13 (KQ1,KQ3) 70 (KQ2,KQ3) 24 (KQ4) 16 (KQ5)	July 2008 to July 2011	3	1 2 3 4- 5	4/6 6/7 4/5 all	1 2 3	2/6 1/7 1/5						x	
Effectiveness of Recombinant Human Growth Hormone (rhGH) in the Treatment of Patients With Cystic Fibrosis (23) Phung (October 2010)[19] [Recombinant human growth hormone in the treatment of patients with cystic fibrosis][20]	April 2010	26(KQ1-2, KQ4, KQ6-7) 50 (KQ3) 3(KQ5)	October 2009 to August 2011	16	1-7	all							x		
Therapies for Children With Autism Spectrum Disorders (26) Warren (April 2011)[21] [A systematic review of early intensive intervention for autism spectrum disorders][22]	May 2010	159	November 2009 to October 2011	15	1 2 3-7	10/14 2/3 all			1 2	4/14 1/3			x		
Pain Management Interventions for Hip Fracture (30) Abou-Setta (May, 2011)[23]	December 2010	98	June 2010 to November 2011	1	1 2-4	7/8 all			1	1/8			x		

CER Title (AHRQ #) Author name (publication date) [Journal publication, if available]	Latest search date for CER (across databases)	# of included studies in CER (total or per KQ)	Period covered by update search	# Of included studies in assessment report	Number of conclusions within the key questions in CER by updating status								Updating priority for the CER		
					KQ#	# Conclusions Up-to-date	KQ#	# Conclusions Probably out of date	KQ#	# Conclusions Possibly out of date	KQ#	# Conclusions Out of date	Low	Medium	High
[Comparative effectiveness of pain management interventions for hip fracture: a systematic review][24]															
Comparative Effectiveness of Traumatic Brain Injury and Depression (25) Guillamondegui (April 2011)[25]	June 2010	115	December 2009 to October 2011	29	1-6	all							x		
Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer (13) Wilt (February 2008)[26] [Association Between Hospital and Surgeon Radical Prostatectomy Volume and Patient Outcomes: A Systematic Review][27]	September 2007	436	March 2007 to March 2012	21	1	11/15			1	2/15	1	2/15			x
											2	1/1			
					3	3/3									
					4	1/3					4	2/3			
Assessed by EPC 2															
Comparative Effectiveness of In-Hospital Use of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care (21) Yank (May 2010)[28] [Systematic review: benefits and harms of in-hospital use of recombinant factor VIIa for off-label indications][29]	August 2009	74	February 2009 to January 2012	15	2	2/3			2	1/3			x		
					3a	8/9			3a	1/9					
					3b	3/4			3b	1/4					
					4b-c	all			4a	1/2					
Comparative Effectiveness of Lipid-Modifying Agents (16) Sharma (September 2009)[30] [Systematic review: comparative effectiveness and harms of combination therapy and monotherapy for dyslipidemia][31]	May 2009	101	November 2008 to October 2011	20	1	3/13					1	10/13			x
					2	34/48			2	14/48					
					3	9/25			3	16/25					
Comparative Effectiveness of Radiofrequency Catheter Ablation for Atrial Fibrillation (15) IP (July 2009)[32]	December 2008	120	June 2008 to September 2011	33	1	all							x		
					2	3/5			2	2/5					
					3	3/4			3	1/4					
					4	all									

CER Title (AHRQ #) Author name (publication date) [Journal publication, if available]	Latest search date for CER (across databases)	# of included studies in CER (total or per KQ)	Period covered by update search	# Of included studies in assessment report	Number of conclusions within the key questions in CER by updating status								Updating priority for the CER		
					KQ#	# Conclusions Up-to-date	KQ#	# Conclusions Probably out of date	KQ#	# Conclusions Possibly out of date	KQ#	# Conclusions Out of date	Low	Medium	High
[Systematic review: comparative effectiveness of radiofrequency catheter ablation for atrial fibrillation][33]															
Comparative effectiveness and safety of radiotherapy treatments for head and neck cancer (20) Samson (May 2010)[34]	September 2009	108	March 2009 to August 2011	7	1	2/3			1	1/3				x	
					2	1/2			2	1/2					
									3	all					
					4	all									
Comparative Effectiveness of Angiotensin Converting Enzyme Inhibitors or Angiotensin II Receptor Blockers Added to Standard Medical Therapy for Treating Stable Ischemic Heart Disease(18) Coleman (October 2009)[35] [Systematic review: comparative effectiveness of angiotensin-converting enzyme inhibitors or angiotensin II-receptor blockers for ischemic heart disease][36]	February 2009	60	August 2008 to November 2011	12					1	all			x		
					2-6	all									
							7	all							
Comparative Effectiveness of Nonoperative and Operative Treatments for Rotator Cuff Tears (22) Sedia (July 2010)[37] [Systematic review: nonoperative and operative treatments for rotator cuff tears][38]	September 2009	137	March 2009 to January 2012	15	1- 6	all							x		
Comparative Effectiveness of Terbutaline Pump for the Prevention of Preterm Birth (35) Gaudet,(September 2011)[39] [Effectiveness of Terbutaline Pump for the Prevention of Preterm Birth. A Systematic Review and Meta-Analysis][40]	April, 2011	14	October 2010 to March 2012	0	1-6	all							x		

KQ=key question; CER=comparative effectiveness review

Discussion

Ideally health care decisions and policy making should be informed and based on the most up to date scientific evidence.[41] Although the importance of updating systematic reviews has been well recognized internationally, there has been relative paucity in research and initiatives towards maintaining the currency of systematic reviews.[2] An earlier, one time assessment identified 4 out of 11 AHRQ published CERs as being sufficiently out-of-date that they should be updated or withdrawn.[7] These results indicated a need for the regular surveillance of AHRQ published CERs assessing their current validity. Therefore, the two EPCs, assisted by another EPC, developed and piloted a system for surveillance and identification of triggers for updating CERs published within the AHRQ EHC Program. This report describes the methodology and preliminary work behind the assessment of the need for updating for 14 AHRQ-funded CERs at least 6 months after their publication.

Our preliminary results indicate that a small proportion of CERs may be in need of urgent updating 1-3 years after their last search date. Of the 14 CERs assessed between June 2011 and June 2012, 9 (64%) were classified as having low priority for updating and 3 CERs (22%) had medium priority for updating. Two CERs (14%) were determined to have high priority for updating.

The implementation of the surveillance assessment program to determine the updating status of published AHRQ CERs has not been without challenges. The assessment of currency and validity of conclusions for each KQ of a CER was based on the totality of information compiled through multiple different sources including the qualitative/quantitative signals, expert opinion, and FDA alerts. Although we used operational and standardized definitions throughout the process to ensure relative consistency in the assessments, human judgment is required to interpret the newly identified evidence in relation to the conclusion of the CER. This judgment is a potential source of variability in the assessment. However, our prior work has shown that in at least one explicit assessment, the inter-rater reliability of these judgments was at least moderate.[14]

An additional challenge was the variability in presentation among the original CERs. Not all CERs presented their executive summaries along with KQs and corresponding conclusions in an identical or even similar manner (e.g., degree of detail, format, or level of summarization). For example, in some CERs, conclusions for each KQ were stratified by the outcome and/or intervention, resulting in multiple conclusions. In other instances, executive summaries were not sufficiently detailed to be able to extract a specific clearly formulated conclusion, and therefore, the reviewers had to probe the entire body of text in the CER. Moreover, some conclusions in KQs were not readily amenable to updating with respect to comparative effectiveness. For example, these conclusions included descriptive information on prevalence of certain risk factors in specific populations.

There were a few instances when experts differed in their opinion or the experts' opinion differed from what was demonstrated in newly identified studies. Such observations may reflect the possibility that experts differ in knowledge of their respective content areas (AHRQ CERs enlist the input from a diverse set of technical experts for exactly this reason) as well as how up-

to-date they are in terms of the emerging new literature.[42] Additionally, for most CERs we received back responses from fewer than 50 percent of the experts originally contacted, although since the experts were not a random sample the implications of this low response rate are not as clear.

To our knowledge, this surveillance assessment project is the first large scale effort that has applied methods to assess the updating status of evidence-based reports (or systematic reviews) in a structured and standardized manner. The application of these methods has proven to be relatively feasible, efficient and at the same time a comprehensive and systematic approach for assessing the need for updating individual CERs across a wide range of health interventions. However, it is premature to generalize our findings to a broader population of CERs or systematic reviews, and more data collection is necessary.

Ideally the results of this surveillance assessment should be electronically linked to the original CER reports and any subsequent journal publications so that readers are advised regarding the CER's (or publication's) updating status and the Agency's assessment of when (and if) any given out-of-date CERs might be updated. We recognize that updating a CER is a complex decision involving competing priorities, resources, and other emergent issues. As such, readers should not view all out-of-date CERs as reaching the same level of priority.

Given these preliminary results, we believe that this approach would potentially help AHRQ and other similar agencies in making informed decisions for prioritizing updating needs across different CERs (or systematic reviews). One of the main future objectives of this program should be to further harmonize and improve the above-described methods in terms of their feasibility, reproducibility, and applicability. The data collection and surveillance over time will allow us to gauge better what is the optimal time period or frequency needed for updating purposes. Among the first 14 CERs only one released within one year of surveillance was classified as anything other than "low" priority for updating. If these preliminary results are confirmed then a one year time frame for surveillance maybe a more efficient interval for regular assessment.

Conclusions

We have established a Surveillance Program that evaluated 14 CERs over the course of 12 months. Regarding the need to update, two were classified as "high" priority, three as "medium" priority, and nine a "low: priority, 9 to 55 months after the last search date of the original CER. We have shown that a program for regular and active surveillance of Comparative Effectiveness Reviews is feasible.

References

1. Whitlock EP, Lopez SA, Chang S, et al. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. J Clin Epidemiol. 2010 May;63(5):491-501. PMID 19540721.

2. Garritty C, Tsertsvadze A, Tricco AC, et al. Updating systematic reviews: an international survey. PLoS One. 2010;5(4):e9914. PMID 20376338.

3. The Cochrane Collaboration. Maintaining your review. 2002. http://www.cochrane-net.org/openlearning/html/mod19-2.htm.

4. French SD, McDonald S, McKenzie JE, et al. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? BMC Med Res Methodol. 2005;5:33. PMID 16225692.

5. Shekelle P, Eccles MP, Grimshaw JM, et al. When should clinical guidelines be updated? BMJ. 2001 Jul 21;323(7305):155-7. PMID 11463690.

6. Shekelle PG, Ortiz E, Rhodes S, et al. Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated? JAMA. 2001 Sep 26;286(12):1461-7. PMID 11572738.

7. Shekelle PG, Newberry SJ, Maglione M, et al. Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005-2009) (Prepared by the Southern California Evidence-based Practice Center). Rockville, MD: Agency for Healthcare Research and Quality; October 2009.

8. Moher D, Tsertsvadze A, Tricco AC, et al. When and how to update systematic reviews. Cochrane Database Syst Rev. 2008(1):MR000023. PMID 18254126.

9. Shojania KG, Sampson M, Ansari MT, et al. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007 Aug 21;147(4):224-33. PMID 17638714.

10. Maclean CH, Issa AM, Newberry SJ, et al. Effects of omega-3 fatty acids on cognitive function with aging, dementia, and neurological diseases. [Review] [5 refs]. Evidence Report: Technology Assessment (Summary). 2005(114):1-3.

11. Maclean CH, Newberry SJ, Mojica WA, et al. Effects of omega-3 fatty acids on cancer. Evid Rep Technol Assess (Summ). 2005 Feb(113):1-4. PMID 15777113.

12. Balk EM, Lichtenstein AH, Chung M, et al. Effects of omega-3 fatty acids on coronary restenosis, intima-media thickness, and exercise tolerance: a systematic review. Atherosclerosis. 2006 Feb;184(2):237-46. PMID 16084516.

13. Shekelle PG, Newberry SJ, Wu H, et al. Identifying Signals for Updating Systematic Reviews: A Comparison of Two Methods. Methods Research Report. AHRQ Publication No. 11-EHC042-EF Agency for Healthcare Research and Quality. Rockville, Maryland: June 2011.

14. Chung M, Newberry SJ, Ansari MT, et al. Two methods provide similar signals for the need to update systematic reviews. J Clin Epidemiol. 2012 Jun;65(6):660-8. PMID 22464414.

15. Bruening W, Schoelles K, Treadwell J, et al. Comparative effectiveness of core-needle and open surgical biopsy for the diagnosis of breast lesions [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); December 2009. Report No.: 10-EHC007-EF; 2009. PM:20704042Comparative Effectiveness Reviews, No. 19.

16. Bruening W, Fontanarosa J, Tipton K, et al. Systematic review: comparative effectiveness of core-needle and open surgical biopsy to diagnose breast lesions. Ann.Intern.Med. 2010;152(4):238-46.

17. Nelson HD, Fu R, Humphrey L, et al. Comparative effectiveness of medications to reduce risk of primary breast cancer in women [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); September 2009. Report No.: 09-EHC028-EF; 2009. PM:20704040AHRQ Comparative Effectiveness Reviews, No. 17.

18. Nelson HD, Fu R, Griffin JC, et al. Systematic review: comparative effectiveness of medications to reduce risk for primary breast cancer. Ann.Intern.Med. 2009;151(10):703-15.

19. Phung OJ, Coleman CI, Baker EL, et al. Effectiveness of recombinant human growth hormone (rhGH) in the treatment of patients with cystic fibrosis [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); October 2010. Report No.: 11-EHC003-EF; 2010. PM:21938797Comparative Effectiveness Reviews, No. 23.

20. Phung OJ, Coleman CI, Baker EL, et al. Recombinant human growth hormone in the treatment of patients with cystic fibrosis. Pediatrics. 2010;126(5):e1211-e26.

21. Warren Z, Veenstra-Vanderweele J, Stone W, et al. Therapies for Children With Autism Spectrum Disorders [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); April 2011. Report No.: 11-EHC029-EF; 2011. PM:21834171Comparative Effectiveness Reviews, No. 26.

22. Warren Z, McPheeters ML, Sathe N, et al. A systematic review of early intensive intervention for autism spectrum disorders. Pediatrics. 2011;127(5):e1303-e11.

23. Abou-Setta AM, Beaupre LA, Jones CA, et al. Pain management interventions for hip fracture [internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); May 2011. Report No.: 11-EHC022-EF; 2011. PM:21938799Comparative Effectiveness Reviews, No. 30.

24. Abou-Setta AM, Beaupre LA, Rashiq S, et al. Comparative effectiveness of pain management interventions for hip fracture: a systematic review. Ann.Intern.Med. 2011;155(4):234-45.

25. Guillamondegui OD, Montgomery SA, Phibbs FT, et al. Traumatic brain injury and depression [Internet] Rockville (MD): Agency for Healthcare Research and Quality (US); April 2011. Report No.: 11-EHC017-EF; 2011. PM:21938798Comparative Effectiveness Reviews, No. 25.

26. Wilt TJ, Shamliyan T, Taylor B, et al. Comparative effectiveness of therapies for clinically localized prostate cancer. Agency for Healthcare Research and Quality (US); 2008. Comparative Effectiveness Reviews, No. 13.

27. Wilt TJ, Shamliyan TA, Taylor BC, et al. Association between hospital and surgeon radical prostatectomy volume and patient outcomes: a systematic review. J Urol. 2008;180(3):820-8.

28. Yank V, Tuohy CV, Logan AC, et al. Comparative Effectiveness of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care. Rockville, MD: Agency for Healthcare Research and Quality. May 2010. Report No.:10-EHC030-EF; 2010. Comparative Effectiveness Review No. 21.

29. Yank V, Tuohy CV, Logan AC, et al. Systematic review: benefits and harms of in-hospital use of recombinant factor VIIa for off-label indications. Ann.Intern.Med. 2011;154(8):529-40.

30. Sharma M, Ansari MT, Soares-Weiser K, et al. Comparative effectiveness of lipid-modifying agents [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); September 2009. Report No.: 09-EHC024-EF; 2009. PM:20704039Comparative Effectiveness Reviews, No. 16.

31. Sharma M, Ansari MT, Abou-Setta AM, et al. Systematic review: comparative effectiveness and harms of combination therapy and monotherapy for dyslipidemia. Ann.Intern.Med. 2009;151(9):622-30.

32. Ip S, Terasawa T, Balk EM, et al. Comparative effectiveness of radiofrequency catheter ablation for atrial fibrillation [Internet] Rockville (MD): Agency for Healthcare Research and Quality (US); July 2009. Report No.: 09-EHC015-EF; 2009. PM:20704038Comparative Effectiveness Reviews, No. 15.

33. Terasawa T, Balk EM, Chung M, et al. Systematic review: comparative effectiveness of radiofrequency catheter ablation for atrial fibrillation. Ann Intern Med. 2009;151(3):191-202.

34. Samson DJ, Ratko TA, Rothenberg BM, et al. Comparative effectiveness and safety of radiotherapy treatments for head and neck cancer [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); May 2010. Report No.: 10-EHC014-EF; 2010. http://www.ncbi.nlm.nih.gov/books/NBK45242/pdf/TOC.pdfComparative Effectiveness Reviews, No. 20.

35. Coleman CI, Baker WL, Kluger J, et al. Comparative effectiveness of angiotensin converting enzyme inhibitors or angiotensin II receptor blockers added to standard medical therapy for treating stable ischemic heart disease [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); October 2009. Report No.: 10-EHC002-EF; 2009.

PM:20704041Comparative Effectiveness Reviews, No. 18.

36. Baker WL, Coleman CI, Kluger J, et al. Systematic review: comparative effectiveness of angiotensin-converting enzyme inhibitors or angiotensin II-receptor blockers for ischemic heart disease. Ann.Intern.Med. 2009;151(12):861-71.

37. Seida JC, Schouten JR, Mousavi SS, et al. Comparative effectiveness of nonoperative and operative treatments for rotator cuff tears [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); July 2010. Report No.: 10-EHC050-EF; 2010. PM:21028756Comparative Effectiveness Reviews, No. 22.

38. Seida JC, LeBlanc C, Schouten JR, et al. Systematic review: nonoperative and operative treatments for rotator cuff tears. Ann.Intern.Med. 2010;153(4):246-55.

39. Gaudet L, Singh K, Weeks L, et al. Terbutaline Pump for the Prevention of Preterm Birth. Agency for Healthcare Research and Quality (US); 2011. Comparative Effectiveness Reviews, No. 35.

40. Gaudet LM, Singh K, Weeks L, et al. Effectiveness of terbutaline pump for the prevention of preterm birth. A systematic review and meta-analysis. PLoS One. 2012;7(2):e31679.

41. Tsertsvadze A, Maglione M, Chou R, et al. Updating comparative effectiveness reviews: current efforts in AHRQ's Effective Health Care Program. J Clin Epidemiol. 2011;64(11):1208-15.

42. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. Ann Intern Med. 2011;154(1):50-5.

Abbreviations

AHRQ	Agency for Healthcare Research and Quality
CER(s)	Comparative Effectiveness Review(s)
CVD	Cardiovascular disease
EPC(s)	Evidence-based Practice Center(s)
FDA	U.S. Food and Drug Administration
MHRA	Medicines and Healthcare products Regulatory Agency
TEP	Technical Expert Panel