Methods Research Report

Framework for Determining Research Gaps During Systematic Review: Evaluation

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Research Report

Framework for Determining Research Gaps During Systematic Review: Evaluation

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

Contract No. 290-2007-10061-I

Prepared by:
Johns Hopkins University Evidence-based Practice Center
Baltimore, MD

Investigators:
Karen A. Robinson, Ph.D.
Oluwaseun Akinyede, M.P.H.
Tania Dutta, M.S., M.P.P.
Veronica Ivey Sawin, B.A.
Tianjing Li, M.D., Ph.D.
Merianne Rose Spencer, B.S.
Charles M. Turkelson, Ph.D.
Christine Weston, Ph.D.

AHRQ Publication No. 13-EHC019-EF
February 2013

This report is based on research conducted by the Johns Hopkins University Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10061-I). 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.

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

Suggested citation: Robinson KA, Akinyede O, Dutta T, Sawin VI, Li T, Spencer MR, Turkelson CM, Weston C. Framework for Determining Research Gaps During Systematic Review: Evaluation. Methods Research Report. (Prepared by Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2007-10061-I.) AHRQ Publication No. 13-EHC019-EF. Rockville, MD: Agency for Health care Research and Quality. February 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

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 email 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

Christine Chang, M.D., M.P.H.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the continuing support of our AHRQ Task Order Officer, Christine Chang, M.D., M.P.H. We express our gratitude to the following individuals for their contributions to this project: Kay Dickersin, Ph.D., Steven N. Goodman, M.D., M.H.S., Ph.D., and Sean Tunis, M.D., M.Sc. We extend our appreciation to the team members from EPCs that evaluated the use of the framework (listed below), all of whom provided thoughtful advice and input.

Participating EPCs

Blue Cross Blue Shield
Johns Hopkins University
McMaster University
Oregon Health and Science University
University of Alberta
University of Minnesota
Vanderbilt University

Peer Reviewers

Jean-Marie Guise, Ph.D.
Oregon Health and Science University
Portland, OR

Melissa McPheeters, M.P.H., Ph.D.
Vanderbilt University
Nashville, TN

Steve Pearson, M.D., M.Sc.
Institute for Clinical and Economic Review
Boston, MA

P. Lina Santaguida, Ph.D.
McMaster University
Hamilton, ON, Canada

Claudia Witt, M.D., M.B.A.
Institute of Social Medicine, Epidemiology and Health Economics
Charite University Medical Center
Berlin, Germany

Framework for Determining Research Gaps During Systematic Review: Evaluation

Structured Abstract

Background. Research gaps prevent systematic reviewers from making conclusions and, ultimately, limit our ability to make informed health care decisions. While there are well-defined methods for conducting a systematic review, there has been no explicit process for the identification of research gaps from systematic reviews. In a prior project we developed a framework to facilitate the systematic identification and characterization of research gaps from systematic reviews. This framework uses elements of PICOS (Population, Intervention, Comparison, Outcomes, Setting) to describe the gaps and categorizes the reasons for the gaps as (A) insufficient or imprecise information, (B) biased information, (C) inconsistent or unknown consistency results, and/or (D) not the right information.

Objective. To further develop and evaluate a framework for the identification and characterization of research gaps from systematic reviews.

Methods. We conducted two types of evaluation: (1) We applied the framework to existing systematic reviews, and (2) Evidence-based Practice Centers (EPCs) applied the framework either during a systematic review or during a future research needs project (FRN). EPCs provided feedback on the framework using an evaluation form.

Results. Our application of the framework to 50 systematic reviews identified about 600 unique research gaps. Key issues emerging from this evaluation included the need to clarify instructions for dealing with multiple comparisons (lumping vs. splitting) and need for guidance on applying the framework retrospectively. We received evaluation forms from seven EPCs. EPCs applied the framework in 8 projects, five of which were FRNs. Challenges identified by the EPCs led to revisions in the instructions including guidance for teams to decide a priori whether to limit the use of the framework to questions for which strength of evidence has been assessed, and the level of detail needed for the characterization of the gaps.

Conclusions. Our team evaluated a revised framework, and developed guidance for its application. A final version is provided that incorporates revisions based on use of the framework across existing systematic reviews and feedback from other EPCs on their use of the framework. Future research is needed to evaluate the relative costs and benefits of using the framework, for review authors and for users of the systematic reviews.

Contents

Introduction

The identification of gaps from systematic reviews is essential to the practice of ‘‘evidence-based research.’’ Health care research should begin and end with a systematic review.[1-3] A comprehensive and explicit consideration of the existing evidence is necessary for the identification and development of an unanswered and answerable question, for the design of a study most likely to answer that question, and for the interpretation of the results of the study.[4]

In a systematic review, the consideration of existing evidence often highlights important areas where deficiencies in information limit our ability to make decisions. We define a research gap as a topic or area for which missing or inadequate information limits the ability of reviewers to reach a conclusion for a given question. A research gap may be further developed, such as through stakeholder engagement in prioritization, into research needs. Research needs are those areas where the gaps in the evidence limit decision making by patients, clinicians, and policy makers. A research gap may not be a research need if filling the gap would not be of use to stakeholders that make decisions in health care. The clear and explicit identification of research gaps is a necessary step in developing a research agenda. Evidence reports produced by Evidence-based Practice Centers (EPCs) have always included a future research section. However, in contrast to the explicit and transparent steps taken in the completion of a systematic review, there has not been a systematic process for the identification of research gaps.

In a prior methods project, our EPC set out to identify and pilot test a framework for the identification of research gaps.[5,6] We searched the literature, conducted an audit of EPC evidence reports, and sought information from other organizations which conduct evidence synthesis. Despite these efforts, we identified little detail or consistency in the frameworks used to determine research gaps within systematic reviews. In general, we found no widespread use or endorsement of a specific formal process or framework for identifying research gaps using systematic reviews.

We developed a framework to systematically identify research gaps from systematic reviews. This framework facilitates the classification of where the current evidence falls short and why the evidence falls short. The framework included two elements: (1) the characterization the gaps and (2) the identification and classification of the reason(s) for the research gap.

The PICOS structure (Population, Intervention, Comparison, Outcome and Setting) was used in this framework to describe questions or parts of questions inadequately addressed by the evidence synthesized in the systematic review. The issue of timing, sometimes included as PICOTS, was considered separately for Intervention, Comparison, and Outcome. The PICOS elements were the only sort of framework we had identified in an audit of existing methods for the identification of gaps used by EPCs and other related organizations (i.e., health technology assessment organizations). We chose to use this structure as it is one familiar to EPCs, and others, in developing questions.

It is not only important to identify research gaps but also to determine how the evidence falls short, in order to maximally inform researchers, policy makers, and funders on the types of questions that need to be addressed and the types of studies needed to address these questions. Thus, the second element of the framework was the classification of the reasons for the existence of a research gap. For each research gap, the reason(s) that most preclude conclusions from being made in the systematic review is chosen by the review team completing the framework. To leverage work already being completed by review teams, we mapped the reasons for research gaps to concepts from commonly used evidence grading systems. Briefly, these categories of reasons, explained in detail in the prior JHU EPC report [5], are:

A. Insufficient or imprecise information
B. Biased information
C. Inconsistent or unknown consistency results
D. Not the right information

The framework facilitates a systematic approach to identifying research gaps and the reasons for those gaps. The identification of where the evidence falls short and how the evidence falls short is essential to the development of important research questions and in providing guidance in how to address these questions.

As part of the previous methods product, we developed a worksheet and instructions to facilitate the use of the framework when completing a systematic review (See Appendix A). Preliminary evaluation of the framework and worksheet was completed by applying the framework to two completed EPC evidence reports. The framework was further refined through peer review. In this current project, we extend our work on this research gaps framework.

Our objective in this project was to complete two types of further evaluation: (1) application of the framework across a larger sample of existing systematic reviews in different topic areas, and (2) implementation of the framework by EPCs. These two objectives were used to evaluate the framework and instructions for usability and to evaluate the application of the framework by others, outside of our EPC, including as part of the process of completing an EPC report. Our overall goal was to produce a revised framework with guidance that could be used by EPCs to explicitly identify research gaps from systematic reviews.

Methods

We completed four steps as outlined below.

Review and Revise Framework and Develop Detailed Instructions

The framework and instructions were reviewed by team members, some of whom were not involved in the initial project. The framework and instructions were modified based on discussion.

Test Framework and Instructions Through Application to Existing Systematic Reviews

We tested the application of the revised framework and instructions with a sample of 50 systematic reviews of randomized controlled trials of clinical topics.

Identification and Selection of Systematic Reviews

We applied the framework to all eligible EPC reports from 2009 to 2011. (Reports from 2007 to 2008 were included in the audit conducted in our prior report). We searched the AHRQ Web site for reports posted from January 1, 2009 to December 12, 2011 (www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports). We retrieved reports for consideration by selecting the heading "Search for Guides, Reviews, and Reports"; selecting, under Report Types, "Research Reviews" and then selecting, under Project Status, "Final."

We also applied the framework to a random sample of Cochrane systematic reviews from 2009 to 2011. We searched The Cochrane Database of Systematic Reviews for reviews published from January 1, 2009, to December 12, 2011. The search was completed by selecting the date range 2009-2011, all issues, and restricting to "reviews."

Search results for the EPC reports and Cochrane reviews were screened serially by two team members using title and abstract to identify systematic reviews that:

- were published or completed within the time range of interest
- represented final or complete reviews
- addressed a clinical topic
- addressed questions about effectiveness or comparative effectiveness of therapies
- included randomized controlled trials

All eligible EPC reports were included. All Cochrane reviews were entered with a corresponding autogenerated reference number into a spreadsheet for random selection. Randomly selected Cochrane reviews were then screened using criteria and process described above. We selected the number of Cochrane reviews that, when added to the included EPC reports, would equal a combined total of 50 systematic reviews.

Application of Framework to Systematic Reviews

Four team members applied the framework to the 50 systematic reviews, as pairs of independent reviewers for each systematic review. Each reviewer had a background in epidemiology and was specifically trained in the use of the framework. To track progress and

maintain the results, the framework worksheet was translated to forms on DistillerSR (EvidencePartners, Ottawa, ON, Canada) and full-text articles of all eligible systematic reviews were uploaded. Pilot testing of the revised framework (from Review and Revise Framework and Develop Detailed Instructions above) was conducted in October and November 2011. A training session on the use of the framework as translated into online forms was held December 9, 2011. Pilot testing of the system in DistillerSR was completed at the end of December 2011, with abstraction starting December 22, 2011. Abstraction was completed by April 1, 2011. Reviewers were asked to track and share any issues encountered in applying the framework. A comparison of the information abstracted by each reviewer was also completed to highlight any discrepancies that might indicate issues to address in the framework or instructions. A third team member reviewed all abstractions and brought forward to the team any apparent discrepancies or issues in the characterization of gaps or the reasons for gaps. These were discussed and common issues identified, for which responses were determined (i.e., revisions to framework or instructions).

Evaluate Implementation of Framework

We issued multiple invitations for the 14 EPCs to apply the framework to identify gaps in one or more of one of their projects. An invitation was issued during presentations at both the spring and fall 2011 EPC Directors' meetings, as well as via email (January 2012). EPCs were informed that any costs for participation in this, as for other methods projects or workgroups, could be covered under a general task order through the EPC program. EPCs agreeing to participate were sent reminders in May, June and July 2012.

An evaluation form was developed (Appendix B) to solicit structured feedback from the EPCs. Open-ended questions requested feedback on specific advantages and challenges encountered in applying the framework. There was no restriction on the type of review or future research needs project (FRN), in terms of question(s) or study design, that the EPC could consider using in applying the framework. (FRNs are projects within the EPC program that engage various stakeholders to develop and prioritize future research needs identified from EPC evidence reviews.) EPCs were asked to submit a completed evaluation form after use of the framework. EPCs were not asked to submit completed framework worksheets.

Revise and Finalize Framework and Instructions

Based on results of the evaluations, our team revised the framework and instructions.

Peer Review and Public Commentary

A draft of this report was reviewed by AHRQ representatives and peer reviewers, and was posted for public view and comment. Comments received were reviewed and a report of comments and their disposition was prepared and submitted with the revised report.

Results

Review and Revise Framework and Develop Detailed Instructions

The team reviewed and discussed the original framework and instructions. Revisions were made iteratively and based on consensus. The initial revised framework and instructions are provided in Appendix C. The primary revision of the framework was the addition of sub-categories for the reasons for the gap. The team felt that further granularity within the categories of reasons for gaps would make completion of the framework more straightforward for review teams, and would ease translation of research gaps to specific research questions, with guidance for studies needed to address those questions.

Definitions for each subcode were added to the instructions.

The specific reasons for gaps are listed in the footnote of the worksheet and described below:

A. Insufficient or imprecise information

Information is insufficient or imprecise if data are sparse and thus uninformative and/or confidence intervals are wide and thus can include conflicting results or conclusions.

A1 – This reason should be selected if no studies are identified.

A2 – This reason should be selected if a limited number of studies are identified.

A3 – This reason should be selected if the sample sizes or event rates in the available studies are too small to allow conclusions.

A4 – This reason should be selected if the estimate of the effect (such as achieved from a meta-analysis) is imprecise. That is, if the width of the confidence interval is such that the conclusion could be for benefit or harm.

B. Information at risk of bias

The aggregate risk of bias is contingent upon the risk of bias of the individual studies.

B1 – This reason should be selected if the study design(s) are inappropriate to address the question of interest (e.g., non-randomized studies for question where randomized studies are more appropriate).

B2 – This reason should be selected if there are major methodological limitations to the available studies leading to high risk of bias or limited internal validity.

C. Inconsistency or unknown consistency

Consistency is the degree to which results from included studies appear to be similar or in concordance.

C1 – This reason should be selected if only one study is identified. If there is only one available study, even if considered a large sample size, the consistency of results is unknown.

C2 – This reason should be selected if the results from available studies are inconsistent. Elements to consider include whether effect sizes vary widely, if the range of effect sizes is wide, limited or no overlap of confidence intervals, and, as appropriate, if statistical tests, such as I^2, indicate heterogeneity.

D. Not the right information

There are a number of reasons why identified studies might not provide the right information to make conclusions about the review question.

D1 – This reason should be selected if the results from studies might not be applicable to the population of interest.

D2 – This reason should be selected if the duration of the interventions and/or comparisons is considered too short.
D3 – This reason should be selected if participants are not followed up for long enough duration in the included studies.
D4 – This reason should be selected if the optimal and/or most important outcomes are not assessed in the included studies. This reason also includes instances where only data on surrogate outcomes are available while data on more clinical and/or patient-important outcomes are needed.
D5 – This reason should be selected if the results from studies might not be applicable to the setting of interest. This would include cases where the interventions assessed in the studies are not applicable or available in setting of interest.

Test Framework and Instructions Through Application to Existing Systematic Reviews

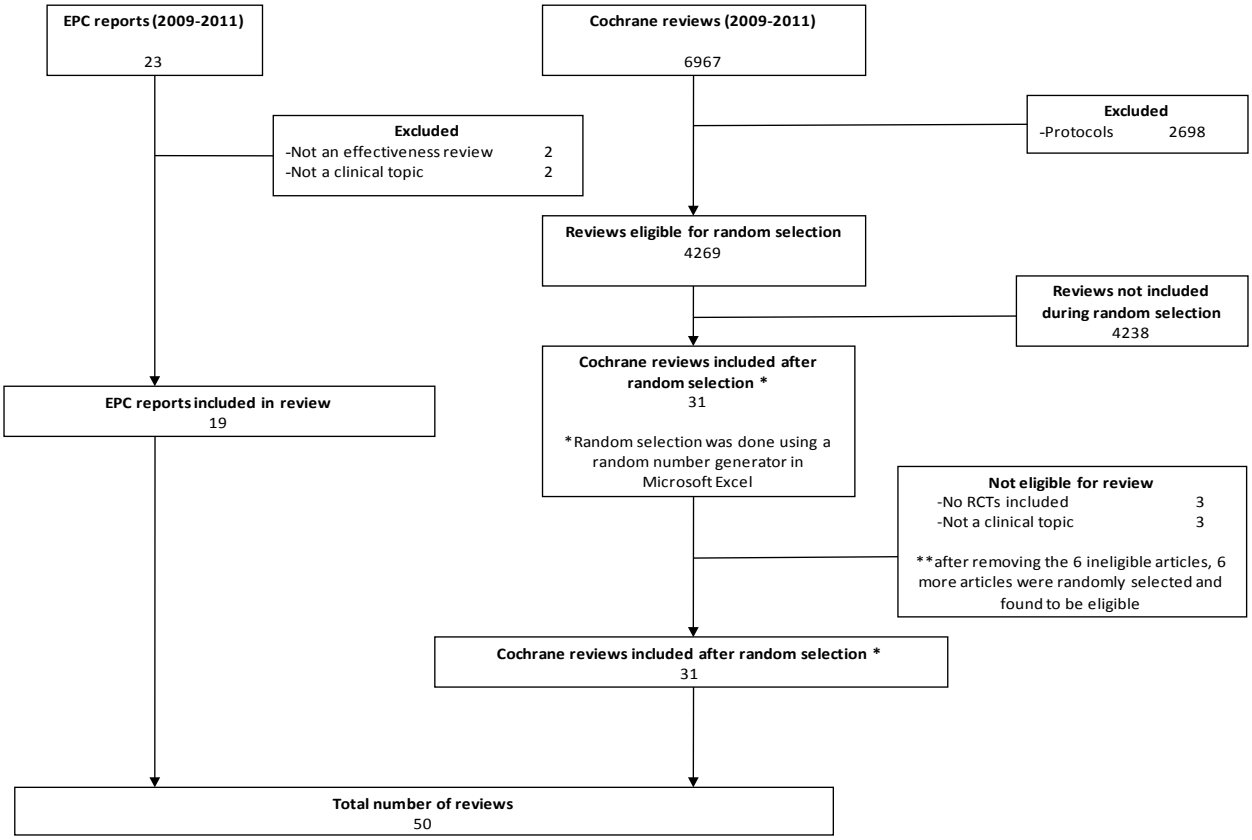
There were 23 EPC reports published on the Effective Health Care Program Web site from January 1, 2009, to December 12, 2011. During screening, four were deemed ineligible due to the following reasons: "not an effectiveness review (n=2) and "not a clinical topic" (n=2).

There were 19 eligible EPC reports; therefore, 31 Cochrane reviews were randomly selected for initial consideration of eligibility criteria to bring the total sample of systematic reviews to 50. There were 6,967 records for January 1, 2009, to December 12, 2011, in the Cochrane Database of Systematic Reviews. Removing protocols, there were 4,269 records. After random sorting and selecting 31 reviews, 6 were determined to be ineligible due to the following reasons: "no RCTs [random controlled trial] included" (n=3) and "not a clinical topic" (n=3). After random selection of an additional six reviews, all six were deemed eligible. A listing of the reviews used in this project is provided in Appendix D. See Figure 1 for a flow diagram of the identification and selection of systematic reviews.

There were 144 review questions included in the 50 systematic reviews. Of the 31 Cochrane reviews, 23 had one review question, 8 had two review questions (average 1.3 questions per review). This was quite different for the EPC reports; the smallest number of review questions was 4 and the highest was 7, with an average of 5.5 review questions per report. The estimated time taken for each reviewer to complete full gaps abstraction was about 7.5 hours for an EPC report and about 3 hours for a Cochrane review. Our four reviewers, two reviewers for each systematic review, took approximately 11 weeks total to complete gaps abstraction for the 50 systematic reviews.

The total number of gaps abstracted, counting those abstracted by each reviewer separately, was 1,830. The number of gaps per Key Question per reviewer ranged from 1 to 165. The average number of gaps abstracted by each reviewer per Key Question was 8.5 (95% confidence interval [CI]: 6.23 to 10.32) and 14.3 (95% CI: 9.80 to 18.87) for the Cochrane reviews and EPC reports respectively. The overall mean number of gaps that each reviewer abstracted per Key Question was 12.7 (95% CI: 9.35 to 16.05).

Figure 1. Identification and selection of reviews



EPC = Evidence-based Practice Center; RCT = randomized controlled trial

However, in reviewing the abstracted information we noted that one reviewer abstracted 165 gaps for one of the questions while the other reviewer abstracted 5 gaps for the same review question. This large discrepancy was due to the former abstractor listing each gap separately and the latter reviewer grouping interventions, comparators and outcomes together. After removing this outlier value, the number of gaps per Key Question per reviewer ranged from 1 to 99. The average number of gaps abstracted by each reviewer per Key Question was 8.5 (95% CI: 6.23 to 10.32) and 12.75 (95% CI: 9.31 to 16.19) for the Cochrane reviews and EPC reports respectively. The overall mean number of gaps that each reviewer abstracted per Key Question was 11.6 (95% CI: 8.94 to 14.07). Based on the former averages, there were about 264 gaps identified from the Cochrane reviews (31 reviews x 8.5 gaps per review) and about 242 gaps identified from the EPC reviews (19 reviews x 12.75 gaps per review). We estimate that if full adjudication were completed there would be about 600 unique research gaps identified.

Insufficient or imprecise information (Gap Reason A) was the most frequent reason that prevented the original systematic reviewers from reaching a conclusion on several research questions (Gap Reason A was used 1,716 times). Inconsistency or unknown consistency among studies (Gap Reason C) was the next common reason for the research gaps (selected 462 times). The reason "not the right information" (Gap Reason D) was chosen 273 times. Biased information (Gap Reason B) was selected 227 times. There were 18 instances where reviewers thought that gaps existed due to another reason (the gap reason did not fit into Gap Reason code

A, B, C, or D). Table 1 provides a breakdown by reason code. Note that multiple reasons could be selected for each gap, and these are total numbers across both reviewers' abstractions.

Table 1. Reasons for gaps

Gap Reason	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Total Number of Times Selected*
A – Insufficient or imprecise information	235	859	304	318	1,716
A-1 No studies	153	315	184	138	790
A-2 Limited number of studies	55	289	62	101	507
A-3 Small sample sizes	19	46	38	37	140
A-4 Imprecise estimate of effect	8	209	20	42	279
B – Biased information	36	71	38	82	227
B-1 Inappropriate study design	9	58	3	21	91
B-2 Major methodological limitations	27	13	35	61	136
C – Inconsistency	3	166	141	152	462
C-1 Consistency unknown	0	113	104	80	297
C-2 Inconsistent results	3	53	37	72	165
D – Not the right information	29	148	67	54	298
D-1 Results not applicable to population	8	34	1	8	51
D-2 Inadequate duration of intervention	11	3	0	7	21
D-3 Inadequate duration of follow-up	5	51	13	4	73
D-4 Most important outcomes not addressed	3	43	52	30	128
D-5 Results not applicable to setting	2	17	1	5	25
Other reason	5	1	5	7	18

*Includes selection by either reviewer; multiple reasons may be selected for a gap

Two trained team members independently applied the framework retrospectively to each existing systematic reviews. A third team member reviewed all abstractions and brought forward to the team apparent discrepancies in the number and type of gaps, as well as the reasons for gaps, abstracted from the same review question. This iterative adjudication process identified a number of issues. The key issues, and our responses, are outlined in Table 2. We did not consider analysis of correlation between the reviewers as necessary or appropriate as we would not expect complete agreement, nor is there a reference standard, for this task. Completing full adjudication was considered beyond the scope of this report, but is planned as future work.

Table 2. Key issues from adjudication process

Issue	Response
After the pilot test, it was apparent that some reviewers were reading through results and determining gaps based on their own interpretation.	Clarified with team the process for this project. Added discussion in instructions about differences and considerations in applying framework during systematic review compared with retrospectively applying framework to existing systematic review.
Some reviewers abstracted details about the population into the worksheet, even when those details were included in the original review question.	We added additional details to the instructions about the elements to be abstracted into PICOS columns of worksheet.
The same gaps were sometimes characterized as one and sometimes treated as separate gaps.	A discussion of the issue of "lumping vs. splitting" has been added to the instructions.
Reviewers put similar text in "Gap Text" or "Notes."	Guidance for how to use these columns has been added to the instructions.

PICOS = population, intervention, comparison, outcomes, setting

Evaluate Implementation of Framework

Of the 14 EPCs, three did not respond to invitations and two declined to participate. Nine EPCs initially agreed to participate and, after several reminders, seven EPCs submitted eight

evaluations (one EPC submitted an evaluation form from two different project teams). Most evaluation forms were submitted in June with the last form submitted July 7, 2012. Detailed results are provided in Appendix E.

Five respondents (63%) used the framework during the completion of a FRN. The remainder applied the framework as part of a systematic review. Because there may be differences in how the framework works when applied retrospectively rather than during a systematic review, we have noted next to the feedback comments whether the framework was applied during a systematic review or as part of a FRN.

All eight respondents indicated that they had previously identified gaps from systematic reviews. However, only one provided a description of methods that had been used by the EPC to identify gaps. The other respondents listed titles of prior FRN topics rather than describe any methods that they had used for the identification of gaps.

Respondents noted a number of advantages to using the framework. The primary advantage noted was that use of the framework facilitated a structured and systematic approach. The structured approach required EPC team members to consider all areas, helped to see areas of redundancies, and kept the team members focused on the scope of the project. Respondents highlighted that the systematic approach was in contrast to the somewhat arbitrary process typically used, and that use of the framework may limit the potential influence of the particular priorities of the research team.

Each respondent provided feedback on the disadvantages and problems, as well as suggestions for the framework and instructions. Some of the issues raised were very similar to those we encountered in applying the framework to the existing systematic reviews. We have provided a detailed response to each comment in Appendix E. We summarize here some of the common issues and our response:

- Implementation of framework to reviews or questions with very limited evidence is cumbersome.

  We agree that the framework may be too specific to use for questions for which, essentially, the entire question is a gap. This may lead to an unmanageable number of gaps and an overly cumbersome process. We revised the instructions to suggest that team members meet prior to the start of the process of identifying gaps to decide on how to handle questions with very limited evidence. We think that in such cases it would still be useful to follow an explicit process, but the framework may be completed for the entire question versus characterizing specific gaps within the question.
- Implementation of framework to questions for which strength of evidence was not assessed was challenging. In other cases, application of framework was replicating work completed in strength of evidence grading.

  We have developed the framework to leverage work already being completed by the EPCs in assessing strength of evidence. However, it was clear from responses that the efficiency of this process is dependent on when the framework is applied and for what specific questions. We had previously suggested that the optimal time for application may be during the writing of the results. We revised the instructions to also include the suggestion that teams consider using the framework for questions, and outcomes that were included in the strength of evidence assessments.
- Completing worksheet when there are gaps comprising multiple comparisons and/or outcomes was cumbersome.

We revised the instructions to include the need to have a discussion and make a decision as to whether to lump or split. For instance, it may be more manageable and useful to abstract gaps by class of intervention or comparison.

Revise and Finalize Framework and Instructions

We added or revised text, and included examples, to provide clarification and further guidance within the instructions. These changes were based on results of the retrospective application of the framework and on feedback from the EPCs, as detailed earlier. The final framework worksheet and instructions are provided in Appendix F.

Discussion

Key Findings

- Initial investigator review of the original framework resulted in the revision of the framework to provide for more specific coding for the reason for the research gap. The team felt that the framework would be easier to apply and also provide more useful information if the categories for reasons for gaps were more detailed.
- Each of the EPC respondents indicated that they had previously identified research gaps from systematic reviews. However, only one described methods used to identify gaps. This finding is in line with results from prior EPC project that EPCs and other systematic reviewers do not use formal methods or frameworks for identifying gaps from systematic reviews.
- Key issues emerged from our application of the framework to existing systematic reviews, and through the evaluation of the use of the framework by EPCs. Common issues included challenges based on what point the framework is applied (during a systematic review or retrospectively using an existing systematic review) and the level of detail needed when characterizing gaps (i.e., lumping versus splitting). We modified the instructions to provide guidance for addressing the challenges, and highlighted areas that should be discussed by team members prior to the process of identifying research gaps using the framework.
- Key advantages to using the framework were also noted by the EPCs. The primary advantage noted was that use of the framework facilitated a structured and systematic approach that helped team members to consider all areas within the scope of the project. The perception of the EPCs, that needs further evaluation, is that the use of the framework provided for a more comprehensive process less open to bias.

Limitations

We chose to apply the framework to 50 systematic reviews to have a number that could be accomplished within our timeframe, yet a large enough number to include systematic reviews across a range of topics. We further limited our application to systematic reviews of randomized controlled trials of clinical topics from two well-known organizations that produce systematic reviews. We imposed this restriction to get a more homogenous set of systematic reviews; to be more certain that differences we saw during the application of the framework were due to potential issues with the framework rather than distinct differences in the study design included, topic addressed by the systematic review (i.e., clinical versus other) or quality of the systematic review. Future testing of the framework, including the use with reviews of other study designs and different sorts of questions, may lead to further revisions of the framework or instructions.

We chose to include Cochrane reviews as these reviews follow a clear and explicit method, and were likely to meet eligibility criteria (i.e., include RCTs and address clinical topic). This was seen as preferable to conducting a search and screen for eligible systematic reviews.

We did not analyze the correlation between the reviewers applying the framework retrospectively as this was not felt to be necessary or appropriate. We would not expect complete agreement, nor is there a reference standard. The process of applying the framework, retrospectively or prospectively, is a task of interpretation and judgment similar to the grading of the strength of evidence. As with the task of grading, appropriate guidance, such as through the

instructions developed, and meeting for the training of and calibration among team members is advised.

We were able to solicit feedback from 8 different EPC teams; however, only 3 of these applied the framework to an ongoing systematic review (and one of these applied the framework after completing of the results section). Further use during a systematic review may identify issues or challenges requiring additional revisions to the framework or instructions.

We did not ask EPCs to track the time it took them to apply the framework. We had discussed this in detail and ultimately felt that it was not a matter of simply completing the framework worksheet. The time to complete the process, similar to grading strength of evidence, is very dependent on the specific review questions, team structure and process, etc. Similarly, it is inherently iterative so it is not clear at what point one would start and stop the completing the worksheet. Issues of how the use of the framework fits into a systematic review project, including considerations of any additional time needed, is an area for future research.

On a related note, we did not assess the best process for application of the framework. We feel that the same team process should be used as in completing the strength of evidence assessments. This would suggest a need for individuals with methodological and domain expertise, but this was not assessed. As with the strength of evidence assessments, there is judgment involved in identifying and characterizing gaps. This suggests a need for team orientation and pilot testing, followed by team discussions after the completion of the process.

While we asked EPCs to try using the framework as part of one of their projects we have limited information about how the EPCs applied the framework. To that end, we don't know if the EPCs applied the framework as an academic exercise (therefore providing information on usability) or if they integrated the completion of the framework with a current project (that might provide us with better idea of usefulness). Similarly, we do not know how, or if, EPCs used the results of applying the framework in their project(s).

Because we do not have a sense of how the use of the framework could fit within the production of an EPC report, including the time needed to complete the process, we cannot make recommendations about the feasibility of using the process. Further, as noted below, future research is needed to assess the potential value of using the framework in order to weigh the potential benefits versus potential costs.

Future Research

There are several outstanding questions or research that may further this work:

- Do the changes made to framework and instructions improve usability? As review teams use the framework there may be additional challenges identified. Further testing across different types of questions, and with reviews including different study designs, may be warranted.
- What is the best process for using the framework? Further evaluation is needed as to whether the process could be conducted by reviewers independently or sequentially. This could include assessments of reliability, specifically whether two reviewers identify the same gaps and reasons for gaps. In addition, a training packet and process could be developed. A set of examples could be provided to illustrate common issues, such as how to use the framework to capture methodological gaps.
- What is the most efficient and appropriate way to integrate this process into the conduct of systematic review or FRN? Is there an optimal time during a systematic review or FRN at which to complete the framework?

- In our previous report, we had proposed a format for presenting research gaps based on the results from the framework.[5] Future research could assess if the use of the framework facilitates the use and presentation of identified gaps. This research could be specific to the different uses of the identified research gaps including (a) to develop FRN sections for systematic reviews, or (b) to solicit input from stakeholders in developing FRN documents.
- Similar to the assessment of strength of evidence, the identification of gaps and the reasons for gaps is based on interpretation and judgment. We outlined in the instructions some issues that should be discussed by a team before starting to identify research gaps. Included are the often arbitrary decisions about which reason(s) is most important in limiting ability to draw conclusions. Future research could determine if a decision system, like a hierarchy, could be established to aid these decisions. Such a ranking might be based on the extent of influence in limiting conclusions and/or the ability to ameliorate the reason(s) through future studies.
- The framework facilitates a more systematic approach to the identification of research gaps, but there is little research on how this information may be utilized and by whom, and whether gaps identified through the framework are more useful. As with other methods of conducting systematic reviews, we think that implementing a more explicit process provides for a more comprehensive product with less bias but, also as with other methods, we don't know if this is true. Does using a formal method to identify gaps, such as the framework, provide value for the systematic review authors and for the users of the systematic review? Is there similar, more or less, benefit when using the framework as part of an FRN project? A comparison to other methods would answer questions such as whether use of the framework identifies more research gaps, whether gaps are characterized more completely, and whether gaps identified in this way provide a more useful basis for the development of research agendas.

As noted earlier, we also plan to adjudicate all of the gaps and reasons for gaps abstracted during this project with a goal of quantitatively and qualitatively describing the characteristics of the gaps, and the relative proportions of research gaps that are due to different types limitations in the evidence. This will provide an evidential basis upon which to improve the design of future RCTs to better address comparative effectiveness questions.

Implications for Practice

We provide here some guidance about the use of the framework, based on the results of this project and our experiences. As noted above, many of the specifics of the integration of the framework within the work of the EPCs represent areas of future research. The first question in determining whether and how to use the framework is determining the purpose of identifying gaps. This will determine the level of granularity needed for the characterization of the research gaps. The second question is related to the systematic review being used to identify gaps. For instance, if the team feels like "the entire systematic review is a gap" then it may not be worthwhile going through the process of using the framework. However, we do feel that even in that case the elements of the framework may help to ensure an explicit process.

We recognize that there are different structures for systematic review teams. We suggest that the framework be applied by the same team members and process as employed in completing the strength of evidence grading, ideally at the time of completing the synthesis and grading. We make this suggestion based on our findings that there are different challenges in applying the

framework retrospectively, and to increase the potential for leveraging the work completed in assessing the strength of evidence.

If completing the identification of research gaps as part of a FRN or otherwise using the framework in a retrospective manner with existing systematic reviews, we suggest the following:

- Restrict abstraction of gaps and reason(s) for gaps to explicit statements made by the review authors. Do not review and interpret the specific results to identify gaps or reasons for gaps. Abstract the gaps and reasons for gaps that are specifically noted by the systematic reviewer authors.
- The team completing the abstraction retrospectively should meet to discuss and agree on sections to be reviewed (text, tables, etc.) as well as what to do if there are apparent discrepancies between sections of the systematic review.
- Inserting the section name and page number (in Notes field of framework worksheet) used to identify a gap might be helpful for adjudication and review.

For an FRN, the gaps identified could be used by the team in developing the list of gaps to be presented to and considered by stakeholders. Depending on the number of gaps identified, the team may choose to prioritize or categorize the gaps prior to presentation to stakeholders.

Whether being completed during a systematic review or applied retrospectively, the instructions (Appendix F) should be reviewed by all participating team members prior to use of the framework. The instructions provide the current guidance for the use of the framework. To leverage the work of assessing strength of evidence, the relevant guidance on the grading system should also be reviewed. Pilot testing should be completed with, as in strength of evidence assessment training, meetings with the full team to calibrate judgments. As noted in the instructions, the research gap framework may be used in different formats (Word, Excel, Access, and DistillerSR) depending on the process being employed by the review team.

Conclusions

In our prior project, we found that very few systematic reviewers used an explicit method to identify research gaps. We completed further evaluation and development of a framework to identify research gaps from systematic reviews. While our focus in this project was on developing the framework for use by EPCs, the framework is not EPC-specific and may be applied by others conducting systematic reviews and/or identifying research gaps from systematic reviews. Future research is needed, especially to evaluate the potential benefit and feasibility of identifying research gaps using the framework. Our framework may be applied during the conduct of or using existing systematic reviews to facilitate an explicit process to characterize where the current evidence falls short and why or how the evidence falls short.

References

1. Clarke M, Alderson P, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. JAMA. 2002; 287(21):2799-801.
2. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? JAMA. 1998; 280(3):280-2.
3. Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med. 2007; 100(4):187-90.
4. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. Ann Intern Med. 2011 Jan 4;154(1):50-5.
5. Robinson KA, Saldanha IJ, Mckoy NA. Frameworks for determining research gaps during systematic reviews. Methods Future Research Needs Report No. 2 AHRQ Publication No. 11-EHC043-EF. Rockville, MD: Agency for Healthcare Research and Quality. June 2011. PMID: 21977524.
6. Robinson KA, Saldanha IJ, Mckoy NA. Development of a framework to identify research gaps from systematic reviews. J Clin Epidemiol. 2011 Dec;64(12):1325-30. PMID: 21937195.

Appendix A. JHU EPC Frameworks Project: Research Gaps Worksheet and Instructions (Original)

JHU EPC Frameworks Project: Research Gaps Worksheet (Original)
<Project Name>
Research Gap Worksheet

Completed by – _____
Date – _____
Page ____ of ____
Key Question Number – _____

Serial no.	Reason(s) for gap*	Population (P)	Intervention (I)	Comparison (C)	Outcomes (O)	Setting (S)	Free text of gap	Notes
Example	B	Women with gestational diabetes	Metformin	Any insulin	Neonatal hypoglycemia, NICU admissions	-		-
Example	D	-	-	-	-		How should the physician assess asthma or bronchodilator responsiveness?	

***** Reasons for Gap -

A. Insufficient or imprecise information
B. Biased information
C. Inconsistency or unknown consistency
D. Not the right information

Instructions for research gaps abstraction worksheet (Original)
Dec 2010

A research gap is a topic or area for which missing or inadequate information limits the ability of reviewers to reach a conclusion on a given question. This worksheet is designed to facilitate the identification and organization of research gaps during evidence reviews sponsored by AHRQ. Our aim was to design a simple, user-friendly worksheet to help investigators record research gaps. We envision that investigators would fill out this worksheet soon after the data synthesis phase, while in the process of writing the results section of the evidence report.

To facilitate the aggregation of research gaps identified by different people, each person should put his/her name/initials and date of completion on the top right corner of the sheet. Each person should also write the worksheet page number and the key question number on the top right corner of the sheet. We encourage members to be consistent in how they choose to fill out this worksheet, both within themselves as well as with other members of the investigative team.

In the worksheet table, each row is one research gap and is numbered accordingly ("Serial Number").

Reason(s) for Gaps

This column allows members to indicate why the research gap exists. The classification of the reasons for gaps are listed and coded in the legend of the gaps abstraction worksheet. Members should choose the most important reason(s) for the existence of the research gap. That reason selected should be the reason(s) that most precludes conclusions from being made. Put another way, members should consider what would be needed to allow for conclusions to be made. Members may choose to enter codes for more than one reason in this column, as appropriate. The specific reasons for gaps are listed in the footnote of the table and described below:

A. <u>Insufficient or imprecise information</u>
   Insufficient information in identified studies can arise if no studies are identified, if a limited number of studies are identified, or if the sample sizes in the available studies are too small to allow conclusions. If the information available in identified studies is insufficient to allow a conclusion or if the estimate of the effect (usually achieved from a meta-analysis) is imprecise there is a research gap.
   Correspondence to grading systems:
   - *EPC SOE:* **Precision** is a required domain.
   - *GRADE*: The GRADE Working Group advises decreasing the grade of the quality of the evidence if the data are "**imprecise** or sparse".
   - *USPSTF:* The following questions are considered while grading the evidence:
     - "How many studies have been conducted that address the key question(s)?"
     - "How large are the studies? (i.e., what is the precision of the evidence?)"

B. Biased information

The aggregate risk of bias is contingent upon the risk of bias of the individual studies.[#19] In addition to considering methodological limitations of studies, the appropriateness of the study design should also be considered.

Correspondence to grading systems:

- *EPC SOE:* **Risk of bias** is a required domain. It incorporates the elements of **study design** and **aggregate quality** of the studies under consideration.
- *GRADE:* **Study quality** and **study design** are key elements.
- *USPSTF:* The following questions are considered while grading the evidence:
  - "To what extent are the existing studies of high quality? (i.e., what is the internal validity?)"
  - "Do the studies have the appropriate research design to answer the key question(s)?"

C. Inconsistency or unknown consistency

Consistency is the degree to which reported effect sizes from included studies appear to go in the same direction. The two elements are whether effect sizes have the same sign (same side of 'no effect') and whether the range of effect sizes is narrow. However, it should be kept in mind that a statistically significant effect size in one study and an effect size whose confidence interval overlaps null in another study do not necessarily constitute inconsistent results. If there is only one available study, even if considered large sample size, the consistency of results is unknown.

Correspondence to grading systems:

- *EPC SOE:* **Consistency** is a required domain.
- *GRADE:* **Consistency** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "How consistent are the results of the studies?"

D. Not the right information

There are a number of reasons why identified studies might not provide the right information. First, results from studies might not be applicable to the population and/or setting of interest. Second, the optimal or most important outcomes might not be assessed. Third, the study duration might be too short and patients might not be followed up for long enough duration to adequately assess some outcomes which might be most important.

Correspondence to grading systems:

- *EPC SOE:* **Directness** is a required domain. It also incorporates the element of surrogate versus clinical outcomes.
- *GRADE:* **Directness** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "To what extent are the results of the studies generalizable to the general US primary care population and situation? (i.e., what is the external validity?)"

Characterization of Research Gaps

To further characterize the research gaps we propose using the PICOS framework using the population (P), intervention (I), comparison (C), outcomes (O), and setting (S). Those elements which are inadequately addressed in the evidence base should be characterized. The other relevant elements will be apparent from the key question from which the research is derived. It follows that for research questions that do not relate to a specific key question, all available elements of the research gap should be characterized.

Population (P) – In this column, team members should be as specific as possible about the age, sex, race/ethnicity, clinical stage, etc. of the population that is not adequately represented in the evidence base. However, it should be recognized that research gaps often do not relate to any specific population but refer to the general population.

Intervention (I) – In this column, team members should specify the name of the intervention that is inadequately included in the evidence base (generic names of drugs and devices are preferred), the duration of the intervention, its dose, its frequency, who will administer it, etc. As with the population, it may not always be appropriate to specify great detail about the intervention.

Comparison (C) – In this column, team members should provide the same relevant details about the comparative intervention as for the intervention of interest – name of comparative intervention, its duration, its dose, its frequency, who will administer it, etc. If the comparison is 'any other intervention', this should be indicated. Similarly, if the comparison is 'no intervention' or placebo, it should be specified as such. It should also be recognized that there may be instances where there is no specific comparison of interest.

Outcomes (O) – In this column, team members should specify the relevant outcomes of interest that are inadequately included in the evidence base. It may be appropriate to organize outcomes by type of outcomes or to only list the types of outcomes (e.g., maternal outcomes and fetal outcomes, liver outcomes, and renal outcomes). If appropriate, the timing of outcome assessments that are missing should be specified. If there are no specific outcomes of interest, this should be indicated.

Setting (S) – In this column, when appropriate, team members should specify the relevant settings for research gaps.

Special Considerations

Research gaps relating to the accuracy of diagnostic tests can be fit into the PICOS framework by considering the diagnostic test under investigation as the intervention (I) and the gold standard test as the comparison (C). Relevant outcomes (O) in this case could include sensitivity and specificity.

Research gaps relating to the benefit of one form (or frequency) of clinical assessment (e.g., monitoring) versus another can be fit into the PICOS framework by considering these clinical assessments as intervention (I) and comparison (C). The comparison in this case could include a

standard form (or frequency) of clinical assessment or no clinical assessment. Relevant outcomes (O) could include clinical outcomes to assess the benefit of the clinical assessment(s).

Research gaps relating to screening tests can be fit into the PICOS framework by considering these tests as intervention (I) and comparison (C). Relevant outcomes (O) could include clinical outcomes to assess the benefit of the screening test(s).

Research gaps which are difficult to characterize into the PICOS framework should be abstracted in free text form. Interventions could potentially include a range of treatment options, order of treatment options, individualization of treatments, etc. These are often gaps for which it is difficult to identify a clear intervention or comparison of interest.
Examples of research questions derived from such research gaps are: “What are the optimal glucose thresholds for medication use in women with gestational diabetes?”; “In what order should patients with cystic fibrosis perform their airway clearance therapies?” and “How should physicians choose an airway clearance therapy for a given patient with cystic fibrosis?”

Appendix B. JHU EPC Framework Evaluation Form

JHU EPC Framework Evaluation Form

1. EPC Name

2. Date Form Completed - Month (mm) ..... |__|__|

3. Date Form Completed - Day (dd**)** ..... |__|__|

4. Date Form Completed - Year (yyyy) ..... |__|__|__|__|

5. During what stage was the evidence gap framework sheet completed?

(Select only one)

- ○ Systematic Review
- ○ Future Research Needs Document
- ○ Other

6. EPC Project Name

7. Who completed the evidence gap framework sheet?

(Select all that apply)

☐ Principal Investigator
☐ Other Investigator
☐ Research Staff Member
☐ Other (list below)

8. Has your center previously identified gaps from systematic reviews?

○ No (go to Q14)
○ Yes (describe below)

9. Previous Gap Identification - Describe

10. Are there **advantages** to using this framework versus how you identified gaps previously?

○ No (go to Q14)

○ Yes (describe below)

11. Advantages of Framework Sheet - Describe

12. Are there **disadvantages** to using this framework versus how you identified gaps previously?

○ No (go to Q14)

○ Yes (describe below)

13. Disadvantages of Framework Sheet - Describe

14. Were there any problems or issues in using the evidence gap framework sheet?

○ No (go to Q16)

○ Yes (describe below)

15. Problems Using Framework Sheet - Describe

16. Do you have any suggestions to improve the efficiency and/or usefulness of the evidence gap framework sheet?

○ No (go to Q18)

○ Yes (describe below)

17. Framework Sheet Suggestions - Describe

18. Do you have any suggestions to improve the instructions for the framework sheet?

○ No (go to Q18

○ Yes (describe below)

19. Instructions Suggestions – Describe

20. General Comments

21. Contact Name (of person completing this evaluation form)

22. Contact E-mail

23. Contact Phone Number

Submit by June 7, 2012:

E-mail: krobin@jhmi.edu
Phone: 410-502-9216
Fax: 410-502-6952
Mail: Johns Hopkins University
1830 E. Monument Street, Room 8068
Baltimore, MD 21287

Appendix C. JHU EPC Frameworks Project: Research Gaps Worksheet and Instructions

JHU EPC Frameworks Project: Research Gaps Worksheet

Systematic Review ID: _____

Completed by – _____

Date – _____

Page ____ of ____

Key Question Number (Enter "99" if outside scope) – _____

Serial No.	Reason(s) for Gap*	Other Reason(s) for Gap	POPULATION (P)	INTERVENTION (I)	COMPARISON (C)	OUTCOMES (O)	SETTING (S)	Free Text Gap	Notes

*** Reasons for Gap**

Insufficient or Imprecise Information → A1=No studies, **A2**=Limited number of studies, **A3**=Sample sizes too small, **A4**=Estimate of effect is imprecise

Information at Risk of Bias → B1=Inappropriate study design, **B2**=Major methodological limitations in studies

Inconsistency or Unknown Consistency → C1=Consistency unknown (only 1 study), **C2**=Inconsistent results across studies

Not the right information → D1=Results not applicable to population of interest, **D2**=Inadequate duration of interventions/comparisons, **D3**=Inadequate duration of follow-up, **D4**=Optimal/most important outcomes not addressed, **D5**=Results not applicable to setting of interest

Instructions for research gaps abstraction worksheet

Oct 2011

Purpose

A research gap is a topic or area for which missing or inadequate information limits the ability of reviewers to reach a conclusion on a given question. This worksheet is designed to facilitate the identification, description and organization of research gaps during evidence reviews sponsored by AHRQ. When completed during the production of an EPC report, investigators would fill out this worksheet soon after the data synthesis phase, while in the process of writing the results section of the evidence report.

Instructions

1. Enter name of EPC report or systematic review project in upper left hand corner.
2. Complete one or more worksheets for each review question (question included in the systematic review). Indicate question number in top right hand corner. (Enter "99" if gap is outside scope)
3. Initial and date each worksheet.
4. Number the worksheets.
5. Enter gaps into the table, per guidance provided below. In the worksheet table, each row is one research gap and is numbered accordingly ("Serial Number").

Guidance for completing table

Reason(s) for Gaps

Enter the reason(s) for the gap in the second column. The classification of the reasons for gaps are listed and coded in the legend of the gaps abstraction worksheet. Choose the most important reason(s) for the existence of the research gap. The reason selected should be the reason(s) that most precludes conclusions from being made about that question. Put another way, consider what would be needed to allow for conclusions to be made. If that particular reason(s) for gap was resolved, could the reviewer draw a conclusion about the question? Codes for more than one reason may be entered in this column, as appropriate. Reasons that cannot be fit within the defined coding system should be listed in the third column titled "Other Reason(s) for Gap".

The reasons for gap are categorized as:

A. Insufficient or imprecise information
B. Information at risk of bias
C. Inconsistency or unknown consistency
D. Not the right information

For each of these categories, the relevant domain or element from the EPC Strength of Evidence, GRADE and USPSTF are listed. Work completed in grading the body of evidence should be used in completing this worksheet. It may be useful to review the most recent guidance about each of these evidence grading systems.

The specific reasons for gaps are listed in the footnote of the table and described below:

A. Insufficient or imprecise information

Information is insufficient or imprecise if data are sparse and thus uninformative and/or confidence intervals are wide and thus can include conflicting results or conclusions.

A1 – This reason should be selected if no studies are identified.

A2 – This reason should be selected if a limited number of studies are identified.

A3 – This reason should be selected if the sample sizes or event rates in the available studies are too small to allow conclusions.

A4 – This reason should be selected if the estimate of the effect (usually achieved from a meta-analysis) is imprecise. That is, if the width of the confidence interval is such that the conclusion could be for benefit or harm.

Correspondence to grading systems:

- *EPC SOE:* **Precision** is a required domain.
- *GRADE*: The GRADE Working Group advises decreasing the grade of the quality of the evidence if the data are "**imprecise** or sparse".
- *USPSTF:* The following questions are considered while grading the evidence:
  - "How many studies have been conducted that address the key question(s)?"
  - "How large are the studies? (i.e., what is the precision of the evidence?)"

B. Information at risk of bias

The aggregate risk of bias is contingent upon the risk of bias of the individual studies.

B1 – This reason should be selected if the study design(s) are inappropriate to address the question of interest.

B2 – This reason should be selected if there are major methodological limitations to the available studies.

Correspondence to grading systems:

- *EPC SOE:* **Risk of bias** is a required domain. It incorporates the elements of **study design** and **aggregate quality** of the studies under consideration.
- *GRADE:* **Study quality** and **study design** are key elements.
- *USPSTF:* The following questions are considered while grading the evidence:
  - "To what extent are the existing studies of high quality? (i.e., what is the internal validity?)"
  - "Do the studies have the appropriate research design to answer the key question(s)?"

C. Inconsistency or unknown consistency

Consistency is the degree to which results from included studies appear to be similar or in concordance.

C1 – This reason should be selected if only one study is identified. If there is only one available study, even if considered a large sample size, the consistency of results is unknown.

C2 – This reason should be selected if the results from available studies are inconsistent. Elements to consider include whether effect sizes vary widely, if the range of effect

sizes is wide, limited or no overlap of confidence intervals, and, as appropriate, if statistical tests, such as I^2, indicate heterogeneity.

Correspondence to grading systems:

- *EPC SOE:* **Consistency** is a required domain.
- *GRADE:* **Consistency** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "How consistent are the results of the studies?"

D. Not the right information

There are a number of reasons why identified studies might not provide the right information.

D1 – This reason should be selected if the results from studies might not be applicable to the population of interest.

D2 – This reason should be selected if the duration of the interventions and/or comparisons is too short.

D3 – This reason should be selected if participants are not followed up for long enough duration in the included studies.

D4 – This reason should be selected if the optimal and/or most important outcomes are not assessed in the included studies. This reason also includes instances where only data on surrogate outcomes are available while data on more clinical and/or patient-important outcomes are needed.

D5 – This reason should be if the results from studies might not be applicable to the setting of interest. This would include interventions not applicable or available in setting of interest.

Correspondence to grading systems:

- *EPC SOE:* **Directness** is a required domain. It also incorporates the element of surrogate versus clinical outcomes.
- *GRADE:* **Directness** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "To what extent are the results of the studies generalizable to the general US primary care population and situation? (i.e., what is the external validity?)"

Characterization of Research Gaps

To further characterize the research gaps we propose using the PICOS framework using the population (P), intervention (I), comparison (C), outcomes (O), and setting (S). Those elements which are inadequately addressed in the evidence base should be characterized. The other relevant elements will be apparent from the key question from which the research is derived. It follows that for research gaps that do not relate to a specific key question, all available elements of the research gap should be characterized.

Population (P) – In this column, specify as much as possible about the age, sex, race/ethnicity, clinical stage, etc. of the population that is not adequately represented in the evidence base. However, it should be recognized that research gaps often do not relate to any specific population but refer to the general population.

Intervention (I) – In this column, specify the name of the intervention that is inadequately included in the evidence base (generic names of drugs and devices are preferred), the duration of the intervention, its dose, its frequency, who will administer it, etc. As with the population, it may not always be appropriate to specify great detail about the intervention.

Comparison (C) – In this column, provide the same relevant details about the comparative intervention as for the intervention of interest – name of comparative intervention, its duration, its dose, its frequency, who will administer it, etc. If the comparison is 'any other intervention', this should be indicated. Similarly, if the comparison is 'no intervention' or placebo, it should be specified as such. It should also be recognized that there may be instances where there is no specific comparison of interest.

Outcomes (O) – In this column, specify the relevant outcomes of interest that are inadequately included in the evidence base. It may be appropriate to organize outcomes by type of outcomes or to only list the types of outcomes (e.g., maternal outcomes and fetal outcomes, liver outcomes, and renal outcomes). If appropriate, the timing of outcome assessments that are missing should be specified. If there are no specific outcomes of interest, this should be indicated.

Setting (S) – In this column, when appropriate, specify the relevant settings or aspect of setting not adequately addressed in evidence base.

Special Considerations

Research gaps relating to the accuracy of diagnostic tests can be fit into the PICOS framework by considering the diagnostic test under investigation as the intervention (I) and the reference standard test as the comparison (C). Relevant outcomes (O) in this case could include sensitivity and specificity.

Research gaps relating to the benefit of one form (or frequency) of clinical assessment (e.g., monitoring) versus another can be fit into the PICOS framework by considering these clinical assessments as intervention (I) and comparison (C). The comparison in this case could include a

standard form (or frequency) of clinical assessment or no clinical assessment. Relevant outcomes (O) could include clinical outcomes to assess the benefit of the clinical assessment(s).

Research gaps relating to screening tests can be fit into the PICOS framework by considering these tests as intervention (I) and comparison (C). Relevant outcomes (O) could include clinical outcomes to assess the benefit of the screening test(s).

Research gaps which are difficult to characterize into the PICOS framework should be abstracted in free text form. Interventions could potentially include a range of treatment options, order of treatment options, individualization of treatments, etc. These are often gaps for which it is difficult to identify a clear intervention or comparison of interest. It may not be possible to translate these gaps into appropriate research questions. Examples of questions derived from such research gaps are: “What are the optimal glucose thresholds for medication use in women with gestational diabetes?”; “In what order should patients with cystic fibrosis perform their airway clearance therapies?” and “How should physicians choose an airway clearance therapy for a given patient with cystic fibrosis?”

Appendix D. Listing of Reviews Included in Retrospective Application of Framework

Reports from Evidence-based Practice Centers (EPCs) Included in Retrospective Application of JHU Research Gaps Framework

1. Abou-Setta AM, Beaupre LA, Jones CA, et al. Pain Management Interventions for Hip Fracture. Comparative Effectiveness Review No. 30. 2011 May. Rockville, MD: Agency for Healthcare Research and Quality (US); AHRQ Publication No. 11-EHC022-EF.
Rec #: 1

2. Balk EM, Moorthy D ,Obadan NO, et al. Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. Comparative Effectiveness Review No. 32. Rockville, MD: Agency for Healthcare Research and Quality (US); 2011 Jul. AHRQ Publication No. 11-EHC052-EF.
Rec #: 2

3. Bennett WL, Wilson LM, Bolen S, et al. Oral Diabetes Medications for Adults With Type 2 Diabetes: An Update. Comparative Effectiveness Review No. 27. Rockville, MD: Agency for Healthcare Research and Quality (US); March 2011. AHRQ Publication No. 11-EHC038-EF.
Rec #: 3

4. Chou R, McDonagh MS, Nakamoto E, et al. Analgesics for Osteoarthritis: An Update of the 2006 Comparative Effectiveness Review. Comparative Effectiveness Review No. 38. Rockville, MD: Agency for Healthcare Research and Quality(US);
2011 Oct. AHRQ Publication No. 11(12)-EHC076-EF.
Rec #: 4

5. 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. Rockville, MD: Agency for Healthcare Research and Quality(US); 2009 Oct.
Rec #: 5

6. Gartlehner G, Hansen RA, Morgan LC, T et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. Comparative Effectiveness Review No. 46. Rockville, MD: Agency for Healthcare Research and Quality(US); 2011 Dec. AHRQ Publication No. 12-EHC012-EF.
Rec #: 6

7. Gaudet L, Singh K, Weeks L, et al. Terbutaline Pump for the Prevention of Preterm Birth. Comparative Effectiveness Review No. 35. Rockville, MD: Agency for Healthcare Research and Quality(US); 2011 Sep. AHRQ Publication No. 11-EHC068-EF.
Rec #: 7

8. Gaynes BN, Lux L, Lloyd S, et al. Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults. Comparative Effectiveness Review No. 33. Agency for Healthcare Research and Quality(US); 2011 Sep. AHRQ Publication No. 11-EHC056-EF.
Rec #: 8

9. Guillamondegui OD, Montgomery SA, Phibbs FT, et al. Traumatic Brain Injury and Depression. Comparative Effectiveness Review No. 25. Agency for Healthcare Research and Quality(US); 2011 Apr. AHRQ Publication No. 11-EHC017-EF.
Rec #: 9

10. Maglione M, Ruelaz Maher A, Hu J, et al. Off-Label Use of Atypical Antipsychotics: An Update. Comparative Effectiveness Review No. 43. Agency for Healthcare Research and Quality(US); 2011 Sep.
Rec #: 10

11. Nelson HD, Fu R, Humphrey L, et al. Comparative Effectiveness of Medications To Reduce Risk of Primary Breast Cancer in Women. Comparative Effectiveness Review No. 17. Agency for Healthcare Research and Quality(US); 2009 Sep.
Rec #: 11

12. Nicholson WK, Wilson LM, Witkop CT, et al. Therapeutic Management, Delivery, and Postpartum Risk Assessment and Screening in Gestational Diabetes. Evidence Report/Technology Assessment No. 162 . Rockville, MD: Agency for Healthcare Research and Quality (US); 2008 Mar. AHRQ Publication No. 08-E004.
Rec #: 12

13. Phung OJ, Coleman CI, Baker EL, et al. Effectiveness of Recombinant Human Growth Hormone (rhGH) in the Treatment of Patients With Cystic Fibrosis. Comparative Effectiveness Review No. 23. Agency for Healthcare Research and Quality(US); 2010 Oct. AHRQ Publication No. 11-EHC003.
Rec #: 13

14. Samson DJ, Ratko TA, Rothenberg BM, et al. Comparative Effectiveness and Safety of Radiotherapy Treatments for Head and Neck Cancer. Comparative Effectiveness Review No. 20. Agency for Healthcare Research and Quality(US); 2010 May.
Rec #: 14

15. Sanders GD, Coeytaux R, Dolor RJ, et al. Angiotensin-Converting Enzyme Inhibitors (ACEIs), Angiotensin II Receptor Antagonists (ARBs), and Direct Renin Inhibitors for Treating Essential Hypertension: An Update. Comparative Effectiveness Review No. 34. Agency for Healthcare Research and Quality(US); 2011 Jun. AHRQ Publication No. 11-EHC063-EF.
Rec #: 15

16. Seida J, Schouten J, Mousavi S, et al. Comparative Effectiveness of Nonoperative and Operative Treatment for Rotator Cuff Tears. Comparative Effectiveness Review No. 22. Agency for Healthcare Research and Quality (US); 2010 Jul. AHRQ Publication No. 10-EHC050.
Rec #: 16

17. Sharma M, Ansari MT, Soares-Weiser K, et al. Comparative Effectiveness of Lipid-Modifying Agents. Comparative Effectiveness Review No. 16. 2009 Sep.
Rec #: 17

18. Warren Z, Veenstra-VanderWeele J, Stone W, et al. Therapies for Children With Autism Spectrum Disorders. Comparative Effectiveness Review No. 26. Agency for Healthcare Research and Quality(US); 2011 Apr. AHRQ Publication No. 11-EHC029-EF.
Rec #: 18

19. Yank V, Tuohy CV, Logan AC, et al. Comparative Effectiveness of Recombinant Factor VIIa for Off-Label Indications vs. Usual Care. Comparative Effectiveness Review No. 21. Agency for Healthcare Research and Quality(US); 2010 May.
Rec #: 19

Cochrane Reviews Included in Retrospective Application of JHU Research Gaps Framework

1. Adams E, Thomson A, Maher C, et al. Mechanical devices for pelvic organ prolapse in women. Cochrane Database Syst Rev. 2004; (2):CD004010.
Rec #: 20

2. Bamigboye A. A. and Smyth R. Interventions for varicose veins and leg oedema in pregnancy. Cochrane Database Syst Rev. 2007; (1):CD001066.
Rec #: 21

3. Costa J, Espirito-Santo C, Borges A, et al. Botulinum toxin type A therapy for cervical dystonia. Cochrane Database Syst Rev. 2005; (1):CD003633.
Rec #: 22

4. Eccleston C, Williams AC. and Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev. 2009; (2):CD007407.
Rec #: 23

5. Escribano J, Balaguer A, Pagone F, et al. Pharmacological interventions for preventing complications in idiopathic hypercalciuria. Cochrane Database Syst Rev. 2009; (1):CD004754.
Rec #: 24

6. Ford AC, Delaney BC, Forman D, et al. Eradication therapy for peptic ulcer disease in Helicobacter pylori positive patients. Cochrane Database Syst Rev. 2006; (2):CD003840.
Rec #: 25

7. Foster G, Taylor SJ, Eldridge SE, et al. Self-management education programmes by lay leaders for people with chronic conditions. Cochrane Database Syst Rev. 2007; (4):CD005108.
Rec #: 26

8. Gagnon AJ, and Sandall J. Individual or group antenatal education for childbirth or parenthood or both. Cochrane Database Syst Rev. 2007; (3):CD002869.
Rec #: 27

9. Hemila H, and Koivula T. T. Vitamin C for preventing and treating tetanus. Cochrane Database Syst Rev. 2008; (2):CD006665.
Rec #: 28

10. Hofmeyr GJ, Kulier R. Abdominal decompression in normal pregnancy. Cochrane Database Syst Rev. 2000; (2):CD001062.
Rec #: 29

11. Jones PW, and Greenstone M. Carbonic anhydrase inhibitors for hypercapnic ventilatory failure in chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2001; (1):CD002881.
Rec #: 30

12. Kelly M, Gillies D, Todd DA, et al. Heated humidification versus heat and moisture exchangers for ventilated adults and children. Cochrane Database Syst Rev. 2010; (4):CD004711.
Rec #: 31

13. Kuschel CA, and Harding JE. Fat supplementation of human milk for promoting growth in preterm infants. Cochrane Database Syst Rev. 2000; (2):CD000341.
Rec #: 32

14. McGuire H, and Hawton K. Interventions for vaginismus. Cochrane Database Syst Rev. 2001; (2):CD001760.
Rec #: 33

15. Moore RA, Derry S, and McQuay HJ. Single dose oral acemetacin for acute postoperative pain in adults. Cochrane Database Syst Rev. 2009; (3):CD007589.
Rec #: 34

16. Nelson R, and Singer M. Primary repair for penetrating colon injuries. Cochrane Database Syst Rev. 2003; (3):CD002247.
Rec #: 35

17. Osiri M, Shea B, Robinson V, et al. Leflunomide for treating rheumatoid arthritis. Cochrane Database Syst Rev. 2003; (1):CD002047.
Rec #: 36

18. Peinemann F, Grouven U, Hemkens LG, et al. Low-dose rate brachytherapy for men with localized prostate cancer. Cochrane Database Syst Rev. 2011; (7):CD008871.
Rec #: 37

19. Petsky HL, Cates CJ, Li A, et al. Tailored interventions based on exhaled nitric oxide versus clinical symptoms for asthma in children and adults. Cochrane Database Syst Rev. 2009; (4):CD006340.
Rec #: 38

20. Piromchai P, Thanaviratananich S, and Laopaiboon M. Systemic antibiotics for chronic rhinosinusitis without nasal polyps in adults. Cochrane Database Syst Rev. 2011; (5):CD008233.
Rec #: 39

21. Sackley C, Disler PB, Turner-Stokes L, et al. Rehabilitation interventions for foot drop in neuromuscular disease. Cochrane Database Syst Rev. 2009; (3):CD003908.
Rec #: 40

22. Sailas E, Fenton M. Seclusion and restraint for people with serious mental illnesses. Cochrane Database Syst Rev. 2000; (2):CD001163.
Rec #: 41

23. Swadpanich U, Lumbiganon P, Prasertcharoensook W, et al. Antenatal lower genital tract infection screening and treatment programs for preventing preterm delivery. Cochrane Database Syst Rev. 2008; (2):CD006178.
Rec #: 42

24. Uman LS, Chambers CT, McGrath PJ, et al. Psychological interventions for needle-related procedural pain and distress in children and adolescents. Cochrane Database Syst Rev. 2006; (4):CD005179.
Rec #: 43

25. van der Schans C, Prasad A, and Main E. Chest physiotherapy compared to no chest physiotherapy for cystic fibrosis. Cochrane Database Syst Rev. 2000; (2):CD001401.
Rec #: 44

26. van Duijvenbode IC, Jellema P, van Poppel MN, et al. Lumbar supports for prevention and treatment of low back pain. Cochrane Database Syst Rev. 2008; (2):CD001823.
Rec #: 45

27. Vemgal P, Ohlsson A. Interventions for non-oliguric hyperkalaemia in preterm neonates. Cochrane Database Syst Rev. 2007; (1):CD005257.
Rec #: 46

28. Villar HC, Saconato H, Valente O, et al. Thyroid hormone replacement for subclinical hypothyroidism. Cochrane Database Syst Rev. 2007; (3):CD003419.
Rec #: 47

29. Wang Y, Pan T, Wang Q, et al. Additional bedtime H2-receptor antagonist for the control of nocturnal gastric acid breakthrough. Cochrane Database Syst Rev. 2009; (4):CD004275.
Rec #: 48

30. Yorke J, Fleming SL, Shuldham CM. Psychological interventions for adults with asthma. Cochrane Database Syst Rev. 2006; (1):CD002982.
Rec #: 49

31. Youssef M. A, Al-Inany HG, Evers JL, et al. Intra-venous fluids for the prevention of severe ovarian hyperstimulation syndrome. Cochrane Database Syst Rev. 2011; (2):CD001302.
Rec #: 50

Appendix E. Detailed Analysis of Evaluation of the Use of the Research Gaps Framework by Evidence-based Practice Centers (EPCs)

Seven EPCs evaluated the Research Gap Framework and submitted 8 evaluation forms (one EPC submitted completed evaluation forms from two different project teams). (Evaluation form is in Appendix B.)

We first provide a summary of the quantitative questions in a table. For each question asking for further details, such as a description of disadvantages, we include the text submitted with the EPCs and projects de-identified. For these questions we have added a column (JHU Response) that includes notes about changes to framework or instructions made in response to the comment(s) or a response, as appropriate. We have also indicated if the form was completed by a team applying the framework during a systematic review (SR) or applying the framework retrospectively during a future research needs project (FRN).

Summary of Responses to Evaluation Form (n, %)

Question Number	Question Text	Number (n=8) No. (%)
Q5	**Stage sheet was completed**	
	Systematic review	**3 (37.5%)**
	Future research needs document	**5 (62.5%)**
	Other	**0 (0%)**
Q7	**Who completed research gap framework worksheet**	
	P.I. only	**4 (50%)**
	Other investigator only	**1 (12.5%)**
	Other investigator and Research staff member	**1(12.5%))**
	Research staff member only	**1(12.5%)**
	Other: team feedback	**1 (12.5%)**
Q8	**EPC had previously completed gap identification**	
	No	**0 (0%)**
	Yes	**8 (100%)**
Q10	**Advantages to using framework vs. previous gap identification method**	
	No	**0 (0%)**
	Yes	**8 (100%)**
Q12	**Disadvantages to using framework vs. previous gap identification method:**	
	No	**1 (12.5%)**
	Yes	**7 (87.5%)**
Q14	**Problems or issues when using framework vs. previous gap identification method:**	
	No	**2 (25%)**
	Yes	**5 (62.5%)**
		1 (12.5%) no answer
Q16	**Suggestions to improve framework sheet efficiency/usefulness:**	
	No	
	Yes	**3 (37.5%)**
		5 (62.5%)
Q18	**Suggestions to improve framework sheet instructions:**	
	No	**5 (62.5%)**
	Yes	**2 (25%)**
		1 (12.5%) no answer

Q9. Describe previous gap identification method

Form	Stage Completed	Description of Gap Identification Method
A	FRN	<XXX>; A Future Research Needs report was also undertaken to systematically prioritize research gaps in the areas of <XXX>, and to develop a list of research questions to address the prioritized gaps based on the systematic review.
B	FRN	Previously, we would review the comparative effectiveness report to determine the number of studies and quality (strength) of evidence to determine the potential research gaps. High quantity + high quality (no gap); high quality + low quantity (no gap);
C	SR	All our reports have a section that identifies gaps. The earlier reviews tended not to be organized around PICOTS.
D	SR	Future Research Needs for the <XXX>.
E	FRN	Have had other Future Research Needs Projects (Different Investigators)
F	FRN	<XXX> Future Research Needs report <XXX> Future Research Needs report Numerous systematic reviews
G	SR	I'm not sure what you mean by "identified gaps from systematic reviews". We regularly write a future research needs section but I'm not sure if these are the same thing.
H	FRN	Yes, this was our third FRN project in addition to the research gaps sections of prior reviews.

Q11. Describe advantages of using framework

Form	Stage Completed	Description of Advantages
A	FRN	Facilitate the use of a systematic process to identify evidence gaps.
B	FRN	This framework provides standardized criteria to identify potential gaps in the literature, which was previously somewhat arbitrary.
C	SR	THE PICOTS framework assists writers in considering all areas.
D	SR	Systematic, transparent way; involvement of different stakeholder groups
E	FRN	The coding / having a list of reasons for the gaps is helpful; but we do not want this to be part of the protocol because we are not sure how to use it and what it adds to the process.
F	FRN	Requires you to be more systematic
G	SR	I can see advantages to using a framework such as this. Without a framework our approach has been fairly non-systematic and may be influenced by priorities of the research team or driven by what they see as the “most important” gaps.
H	FRN	Yes – a structured approach is helpful for constraining the content to researchable topics. It also helps to see where there are redundancies, and helps keep the research team focused on the scope of the project.

Q13. Describe disadvantages of using framework

Form	Stage Completed	Description of Disadvantages	JHU Response
A	FRN	_____	_____
B	FRN	We found that applying this framework to all potential combinations of PICOS for the <XXX> FRN project yielded more than 1000 research gaps. This was due to the large number of populations, settings, and intervention/comparisons and the overall poor quality of the existing literature. The framework is much more practical when there is a manageable number of potential PICOS combinations (e.g. FRN for <XXX> project).	We have added text to the instructions suggesting that teams discuss prior to the use of the framework whether to, and how to, lump or split. For instance, it may be more manageable to abstract gaps by class of intervention and comparison.
C	SR	The overlap with GRADE is less helpful as it is not clear how the gap will assist in the judgments of the SOE. For example a gap in research design….not all issues can be addressed with trials….so not sure how to make this link. Also I think the list of reasons for developing a gap should be expanded. I found I used B2 very often to provide a reason for the gap….and the recommendations are widely varied. Thus some categories are not discriminating enough.	We did not consider the process of identifying gaps as a way to assist in making judgments about SOE. The framework was designed to leverage work completed, if SOE was assessed. We do not see how B2 could be made more specific. It is to be selected if the body of evidence was considered at high risk of bias (this may be for a number of reasons, but is aggregated across the studies). Use of the framework does not preclude providing more details.
D	SR	Involvement of different stakeholder groups may be not representative; information about ongoing studies may be incomplete ; no full representation of the NIH, other funding agencies; the role of industry is unclear	We have clarified at the beginning of the instructions the purpose for the framework – to identify and characterize gaps from systematic reviews. How to solicit stakeholder involvement and prioritize gaps is beyond the scope of this work.
E	FRN	The gaps are not clearly conveyed by the table. The statement in the instructions that “other elements will be apparent from the key question” does not seem to be accurate to us. At the FRN point it is almost too late; too difficult to use. It may have been more helpful during the CER.	We have added some discussion of this, including examples, under Characterization of Research Gaps in the instructions. We agree that there are different challenges in applying the framework retrospectively

Form	Stage Completed	Description of Disadvantages	JHU Response
			versus while completing a systematic review. We have added text in instructions providing some suggestions about how to proceed if doing so retrospectively.
F	FRN	Not all gaps are equally important, so it is not an efficient use of time to be required to complete this chart for every gap. Suggest that only the critical research gaps be prioritized for the chart. It may be too cumbersome for readers to understand. Many of the codes may need to be listed for each gap. It is not clear that using the codes, as opposed to a narrative description, will make the gaps easier or more efficient to understand.	It is unclear how one would determine the 'most critical' gaps without first systematically identifying and characterizing the gaps, such as through use of this framework or other method. The codes and worksheet were developed to aid in abstraction. The future research needs section of the systematic review, or future research needs document, would present the gaps. Our previous report provided a suggested presentation format (also another EPC has produced a report on how to present gaps).
G	SR	The key disadvantage I see is that it may replicate work already done. This may be less of an issue if it was done alongside preparing the results, as was suggested in your instructions. I did it after the review was complete so I found it fairly redundant, as much of this information was already in SOE or summary tables. I'm not sure that it highlighted any issues that were not already known, i.e., very few studies providing data for the same comparisons & outcomes. So it could add a lot of work without providing much additional insight.	We would hope that a team could leverage the work done in completing SOE but we take your point that it could also be redundant. We think this will depend on the team, the specific review, and the timing of applying the framework.
H	FRN	To some degree it can be overly constraining and it really doesn't work well for a review topic on which there is very little available. In this case, the overwhelming gap is that much more research needs to be done, period. Trying to specify at the level of the framework is not yet possible or appropriate. Also, the framework is not ideal for methodologic issues.	We agree that the framework may be too granular to use for questions for which, essentially, the entire question is a gap. We have added some text about these sorts of decisions to the instructions

Q15. Describe problems or issues faced when using framework

Form	Stage Completed	Description of Problems or Issues	JHU Response
A	FRN	_____	_____
B	FRN	The only problem is the same as the disadvantage.	See above.
C	SR	I wasn't sure what would be helpful to you in the free text and notes.	We have added text addressing these sections on the worksheet.
D	SR	Complicated, does not address strength of existing evidence.	We feel that the strength of existing evidence is explicitly considered in the reasons for gaps. Further, we have tried to link the reasons for gaps with the various domains used in different systems to rate the strength of existing evidence.
E	FRN	Not sure the table format adds much value to the process. Seems like we would have to shoe horn items into the table and get little added value from the exercise. Not sure how to complete the PICOTS sections for the types of gaps we identified.	We are not sure of current process used by this EPC team (Q9), so do not have a basis for responding to how use of the framework to identify research gaps in a systematic manner might add value. We have added text to the instructions to clarify characterization of research gaps using PICOS elements in worksheet.
F	FRN	_____	_____

Form	Stage Completed	Description of Problems or Issues	JHU Response
G	SR	For the review I used, we had many, many comparisons (22 drug-drug comparisons for two different conditions within each of 5 key questions) and many outcomes within each of the questions; for many comparisons and outcomes there were very few studies. Therefore, I found the framework rather cumbersome to use. The other challenge was when the outcomes weren't graded. Within the review I used, we only graded outcomes for 2 of the key questions, so for the outcomes (which were numerous) within the other key questions, we had no SOE assessments. So then the reason for gaps was usually A1 (no studies) or A2 (limited number of studies).	We have added to the instructions a discussion of lumping/splitting which, we think, would help in the situation described. We added to instructions decision about whether to review all questions and outcomes, even if not 'graded'.
H	FRN	See question above – it worked for the review in question (<XXX>), but not for another review (<XXX>) that started from all insufficient.	We agree that the framework may be too granular to use for questions for which, essentially, the entire question is a gap. We have added some text about these sorts of decisions to the instructions.

Q17. Suggestions for improving usefulness and efficiency of framework

Form	Stage Completed	Suggestions to Improve Framework	JHU Response
A	FRN	_____	_____
B	FRN	This framework is designed very well for specific projects that contain a manageable number of research gaps. However, for the <XXX> FRN project where literally every combination was determined to be a research gap. It would be impossible to ask expert Stakeholders to evaluate such a large number of research gaps and to then build consensus on prioritization. As exemplified in this project, I do not believe this tool is appropriate for use in all FRN projects and use should be determined on a case-by-case basis by discussion between the investigative team and the TOO.	We have clarified at the beginning of instructions where we envision this framework fitting within the work of a systematic review and future research needs project. We have added text describing decisions to be made about which questions and outcomes to consider (only those assessed for strength of evidence?) and dealing with multiple interventions/comparisons (lumping versus splitting).
C	SR	Might be easier to complete in an excel sheet where some responses can be constrained.	We agree. We completed abstraction for this project using forms in Distiller. We have added a note about this option in the instructions.
D	SR	Research and development framework (used by industry) can be applied using complete information about completed and ongoing studies ; electronic surveys of the representative groups of sponsoring organizations; policy makers, researchers, and consumers (NO “patients”); survey should address group specific interests (implications for funding, research methodology, policy, consumer interests)	We think these comments relate to other aspects of developing a research agenda and are beyond the scope of this project.
E	FRN	If we are going to have a table, it might be more useful to state the gap, then code the reason and the PICOTS issues. I am not sure how the table format is supposed to aid in either making conclusions or communicating them to readers. What is most important? Having gaps with the same reason? Having gaps related to a PICOTS element? Is the table supposed to help you summarize across gaps?	The worksheet was designed to aid in identification of research gaps. The future research needs section of the systematic review, or future research needs document, would present the gaps. Our previous report provided a

Form	Stage Completed	Suggestions to Improve Framework	JHU Response
			suggested presentation format (also another EPC has produced a report on how to present gaps).
F	FRN	Do not think it is necessary to have a separate chart for each key question Instead of "serial number" suggest calling it "gap number"	The gaps are abstracted by question, and characterized by listing the elements of PICOS from the question where evidence is inadequate. Because of this explicit link to questions, each review question should have a worksheet. The alternative is to use the PICOS to flesh out the entire research question needed to address the gap. We have changed the column header to "gap number".
G	SR	As I have alluded to above, it will likely be most efficient to incorporate it at an early stage in the review. It may also be most efficient to focus on some key comparisons or questions or outcomes. Since you mention using the SOE information, should it be based on or driven by the "graded" outcomes?	We have added text about applying the framework retrospectively versus during completion of a systematic review. We have added text to instructions suggesting team make decision about whether to limit consideration and abstraction of gaps to those questions and outcomes that were assessed for strength of evidence.
H	FRN	_____	_____

Q19. Suggestions for improving framework instructions

Form	Stage Completed	Suggestions for Improving Instructions	JHU Response
A	FRN	_____	_____
B	FRN	The instructions were extremely clear.	Thank you.
C	SR	_____	_____
D	SR	Depends on the changes in the research and development framework	
E	FRN	Provide one or two examples of a completed table. Describe how the table can or should be used and clarify what the purpose is.	We have provided, embedded in instructions, some examples to illustrate specific points. We have appended to end of instructions an example of a completed sheet.
F	FRN	_____	_____
G	SR	I found the instructions clear. As I mentioned above, many of the comparisons and outcomes were not graded, therefore the instructions "Work completed in grading the body of evidence should be used in completing this worksheet" are not relevant. What do we do when grading has not been done?	We have revised the text in this section to address this question.
H	FRN	_____	_____

Appendix F. JHU EPC Frameworks Project: Research Gaps Worksheet and Instructions (Final)

JHU EPC Frameworks Project: Research Gaps Worksheet (Final)

Systematic Review ID: _____

Completed by – _____

Date – _____

Page ____ of ____

Key Question Number (Enter "99" if outside scope) – _____

Gap No.	Reason(s) for Gap*	Other Reason(s) for Gap	POPULATION (P)	INTERVENTION (I)	COMPARISON (C)	OUTCOMES (O)	SETTING (S)	Free Text Gap	Notes

*** Reasons for Gap**

Insufficient or Imprecise Information → A1=No studies, **A2**=Limited number of studies, **A3**=Sample sizes too small, **A4**=Estimate of effect is imprecise

Information at Risk of Bias → B1=Inappropriate study design, **B2**=Major methodological limitations in studies

Inconsistency or Unknown Consistency → C1=Consistency unknown (only 1 study), **C2**=Inconsistent results across studies

Not the right information → D1=Results not applicable to population of interest, **D2**=Inadequate duration of interventions/comparisons, **D3**=Inadequate duration of follow-up, **D4**=Optimal/most important outcomes not addressed, **D5**=Results not applicable to setting of interest

Instructions for research gaps framework (Final)
23 July 2012

Purpose

A *research gap* is a topic or area for which missing or inadequate information limits the ability of reviewers to reach a conclusion on a given question.

The *framework*, and accompanying worksheet, is designed to facilitate an explicit process for the identification, description and organization of research gaps during systematic reviews.

When completed during the completion of a systematic review, we suggest that review authors fill out this worksheet soon after the data synthesis phase, while in the process of writing the results section. The results would be used by the team in developing the future research needs section of the report of the systematic review.

The framework may also be applied retrospectively, that is, to identify and characterize gaps from an existing systematic review. For instance, within the Evidence-based Practice Center (EPC) program, the framework may be completed at the start of a Future Research Needs (FRN) project using an existing systematic review that may or may not have been completed by the same team. When completing the framework retrospectively, restrict abstraction of gaps and reason(s) for gaps to explicit statements made by the review authors. Do not review and interpret the specific results to identify gaps or reasons for gaps. Abstract the gaps and reasons for gaps that are specifically noted by the systematic reviewer authors. The team completing the abstraction retrospectively should meet to discuss and agree on sections to be reviewed (text, tables, etc.) as well as what to do if there are apparent discrepancies between sections of the systematic review. Inserting the section name and page number(s) (in Notes field of framework worksheet) used to identify a gap might be helpful for adjudication and review. For an FRN, the gaps identified would be used by the team in developing the list of gaps to be presented to and considered by stakeholders (i.e., gaps may be prioritized or categorized prior to presentation to stakeholders).

There are a number of decisions that a team using the framework should discuss prior to starting the gap identification process. The decisions will be influenced by the purpose for the identification of the gaps:

- At what point will the framework be applied: during completion of a systematic review or retrospectively? We have included guidance for different stages but suggest the optimal time of use is during the writing of the results section of a systematic review.
- Will all questions and outcomes be reviewed for gaps? The team could decide to limit identification to those questions and outcomes for which strength of evidence was assessed (see page 2).
- What level of granularity is needed for the characterization of the gaps? The team should discuss whether to lump or split concepts and, if lumping, how that would be done (see page 6).

Instructions for use of worksheet

1. As required, enter the name of EPC report or systematic review project in upper left hand corner.
2. Complete one or more worksheets for each review question (question included in the systematic review). Indicate question number in top right hand corner. (Enter "99" if gap is outside scope of original systematic review questions.)
3. Initial and date each worksheet.
4. Number the worksheets.
5. Enter gaps into the worksheet, per guidance provided below. In the worksheet table, each row is one research gap and is numbered accordingly ("Gap Number").

Note: The worksheet is provided as a word processing document but it may be translated for use on web-based systems, databases, or spreadsheets.

Guidance for completing worksheet

Coding for the Reason(s) for Research Gaps

Enter the reason(s) for the gap in the second column. The classification of the reasons for gaps are listed and coded in the footnote of the gaps abstraction worksheet. Choose the most important reason(s) for the existence of the research gap. The reason selected should be the reason(s) that most precludes conclusions from being made about that question. In other words, consider what would be needed to allow for conclusions to be made. If that particular reason(s) for gap was resolved, could the reviewer draw a conclusion about the question? Codes for more than one reason may be entered in this column, as appropriate. Reasons that cannot be fit within the defined coding system should be listed in the third column titled "Other Reason(s) for Gap".

The reasons for gap are categorized as:

A. Insufficient or imprecise information
B. Information at risk of bias
C. Inconsistency or unknown consistency
D. Not the right information

For each of these categories, the relevant domain or element from the EPC Strength of Evidence (SOE), GRADE and USPSTF grading systems are listed. It may be useful to review the most recent guidance about each of these evidence grading systems. Work completed in grading the body of evidence may be leveraged in completing this worksheet. The concepts discussed below should be considered and applied in cases where SOE was not assessed. Decide before starting process if all questions and outcomes will be reviewed for gap identification, or only those which were considered in a strength of evidence assessment.

The specific reasons for gaps are listed in the footnote of the worksheet and described below:

A. Insufficient or imprecise information

Information is insufficient or imprecise if data are sparse and thus uninformative and/or confidence intervals are wide and thus can include conflicting results or conclusions.

A1 – This reason should be selected if no studies are identified.

A2 – This reason should be selected if a limited number of studies are identified.

A3 – This reason should be selected if the sample sizes or event rates in the available studies are too small to allow conclusions.

A4 – This reason should be selected if the estimate of the effect (such as achieved from a meta-analysis) is imprecise. That is, if the width of the confidence interval is such that the conclusion could be for benefit or harm.

Note: It would be inconsistent to choose Reason A1 (no studies) and Reason A2 (a limited number of studies) to describe the same gap, since only one or the other can be true. Likewise, Reasons A3 and A4 would not occur at same time as Reason A1.

Correspondence to grading systems:

- *EPC SOE:* **Precision** is a required domain.
- *GRADE*: The GRADE Working Group advises decreasing the grade of the quality of the evidence if the data are "**imprecise** or sparse".
- *USPSTF:* The following questions are considered while grading the evidence:
  - "How many studies have been conducted that address the key question(s)?"
  - "How large are the studies? (i.e., what is the precision of the evidence?)"

B. Information at risk of bias

The aggregate risk of bias is contingent upon the risk of bias of the individual studies.

B1 – This reason should be selected if the study design(s) are inappropriate to address the question of interest (e.g., non-randomized studies for question where randomized studies are more appropriate).

B2 – This reason should be selected if there are major methodological limitations to the available studies leading to high risk of bias or limited internal validity.

Correspondence to grading systems:

- *EPC SOE:* **Risk of bias** is a required domain. It incorporates the elements of **study design** and **aggregate quality** of the studies under consideration.
- *GRADE:* **Study quality** and **study design** are key elements.
- *USPSTF:* The following questions are considered while grading the evidence:
  - "To what extent are the existing studies of high quality? (i.e., what is the internal validity?)"
  - "Do the studies have the appropriate research design to answer the key question(s)?"

C. Inconsistency or unknown consistency

Consistency is the degree to which results from included studies appear to be similar or in concordance.

C1 – This reason should be selected if only one study is identified. If there is only one available study, even if considered a large sample size, the consistency of results is unknown.

C2 – This reason should be selected if the results from available studies are inconsistent. Elements to consider include whether effect sizes vary widely, if the range of effect sizes is wide, limited or no overlap of confidence intervals, and, as appropriate, if statistical tests, such as I^2, indicate heterogeneity.

Note: It would be inconsistent to choose Reason C1 and Reason C2 to describe the same gap, since only one or the other of these reasons can be true.

Correspondence to grading systems:

- *EPC SOE:* **Consistency** is a required domain.
- *GRADE:* **Consistency** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "How consistent are the results of the studies?"

D. Not the right information

There are a number of reasons why identified studies might not provide the right information to make conclusions about the review question.

D1 – This reason should be selected if the results from studies might not be applicable to the population of interest.

D2 – This reason should be selected if the duration of the interventions and/or comparisons is considered too short.

D3 – This reason should be selected if participants are not followed up for long enough duration in the included studies.

D4 – This reason should be selected if the optimal and/or most important outcomes are not assessed in the included studies. This reason also includes instances where only data on surrogate outcomes are available while data on more clinical and/or patient-important outcomes are needed.

D5 – This reason should be selected if the results from studies might not be applicable to the setting of interest. This would include cases where the interventions assessed in the studies are not applicable or available in setting of interest.

Correspondence to grading systems:

- *EPC SOE:* **Directness** is a required domain. It also incorporates the element of surrogate versus clinical outcomes.
- *GRADE:* **Directness** is a key element.
- *USPSTF:* The following question is considered while grading the evidence:
  - "To what extent are the results of the studies generalizable to the general US primary care population and situation? (i.e., what is the external validity?)"

Characterization of Research Gaps

To further characterize the research gaps we propose using the PICOS framework using the population (P), intervention (I), comparison (C), outcomes (O), and setting (S). Those elements of the original review question which are inadequately addressed in the evidence base should be characterized. The other relevant elements will be apparent from the review question from which the research is derived. For research gaps that do not relate to a specific key question, all elements of the research gap should be characterized.

Population (P) – In this column, specify as much as possible about the age, sex, race/ethnicity, clinical stage, etc. of the population that is not adequately represented in the evidence base. However, research gaps often do not relate to any specific population but refer to the general population as outlined in the review question. In that case, it is not necessary to reiterate the population already described in the review question. For example, if the population described by the Key Question is 'pregnant women', there is no need to write 'pregnant women' in this column. This column is designed for other populations, aspects of populations, or subgroups that have not been adequately addressed by the evidence. If the population being studied was 'pregnant women', but none of the studies included pregnant teenagers, or pregnant women over the age of 45, or minority pregnant women, or pregnant women in underdeveloped countries—and the authors of the review consider this a gap—this information would be recorded in this column.

Intervention (I) – In this column, specify the name of the intervention that is inadequately included in the evidence base (generic names of drugs and devices are typically preferred), the duration of the intervention, its dose, its frequency, who will administer it, etc., as appropriate. As for the population, it may not always be appropriate to specify great detail about the intervention.

Comparison (C) – In this column, provide the same relevant details about the comparative intervention as for the intervention of interest – name of comparative intervention, its duration, its dose, its frequency, who will administer it, etc. If the comparison is 'any other intervention', this should be indicated. Similarly, if the comparison is 'no intervention' or placebo, it should be specified as such. It should also be recognized that there may be instances where there is no specific comparison of interest.

Outcomes (O) – In this column, specify the relevant outcomes of interest that are inadequately included in the evidence base. It may be appropriate to organize outcomes by type of outcomes or to only list the types of outcomes (e.g., maternal outcomes and fetal outcomes, liver outcomes, and renal outcomes). If appropriate, the timing of outcome assessments that are missing should be specified. If there are no specific outcomes of interest, this should be indicated.

Setting (S) – In this column, when appropriate, specify the relevant settings or aspect of setting not adequately addressed in evidence base.

Special Considerations

Research gaps relating to the accuracy of diagnostic tests can be fit into the PICOS framework by considering the diagnostic test under investigation as the intervention (I) and the reference standard test as the comparison (C). Relevant outcomes (O) in this case could include sensitivity and specificity.

Research gaps relating to the benefit of one form (or frequency) of clinical assessment (e.g., monitoring) versus another can be fit into the PICOS framework by considering these clinical assessments as intervention (I) and comparison (C). The comparison in this case could include a standard form (or frequency) of clinical assessment or no clinical assessment. Relevant outcomes (O) could include clinical outcomes to assess the benefit of the clinical assessment(s).

Research gaps relating to screening tests can be fit into the PICOS framework by considering these tests as intervention (I) and comparison (C). Relevant outcomes (O) could include clinical outcomes to assess the benefit of the screening test(s).

Free Text Gap column

The Free Text Gap column may be used to characterize research gaps which are difficult to characterize using the PICOS framework. Interventions could potentially include a range of treatment options, order of treatment options, individualization of treatments, etc. These are often gaps for which it is difficult to identify a clear intervention or comparison of interest. It may not be possible to translate these gaps into appropriate research questions. Examples of questions derived from such research gaps are: “What are the optimal glucose thresholds for medication use in women with gestational diabetes?”; “In what order should patients with cystic fibrosis perform their airway clearance therapies?” and “How should physicians choose an airway clearance therapy for a given patient with cystic fibrosis?”

Lumping Versus Splitting

A decision should be made prior to starting abstraction of gaps, either as part of a systematic review or retrospectively, as to how to deal with cases where there is little or no evidence across a broad question resulting in a very high number of comparisons and outcomes with gaps. For example, in such cases, a team may choose to lump together interventions of a certain type or simply note that the entire question is a gap. The team should also discuss and decide *a priori* whether gaps will generally be lumped or split, such as by classes of interventions or types of outcomes. This may depend on the type of question(s) addressed in the review and the purpose for identifying gaps (i.e., the need for granularity). Different decisions could be made for each key question within a systematic review.

Example:

Option 1—Lumping or pooling outcomes with same reason for gap

Reason for gap	P	I	C	O	S
C2		Cognitive behavioral therapy		Pain, mood, and disability	

Option 2—Splitting or separating outcomes with same reason for gap

Reason for gap	P	I	C	O	S
C2		Cognitive behavioral therapy		Pain	
C2		Cognitive behavioral therapy		Mood	
C2		Cognitive behavioral therapy		Disability	

EXAMPLE completed worksheet

Research Gap Worksheet

Project Name: Oral diabetes meds

Completed by – __KR_____
Date – __20 July 2012_____
Page 1 of 1
Key Question Number – __3__

Serial No.	Reason(s) for Gap*	Other Reason(s) for Gap	POPULATION (P)	INTERVENTION (I)	COMPARISON (C)	OUTCOMES (O)	SETTING (S)	Free Tax Gap	Notes
Example	B1			Metformin	Metformin + Any insulin	Weight, lipoproteins	-		-
Example	D1		African-American adults						
Example	A3, A4			Sulfonylurea	GLP-1 agonist	HDL			
Example	D1, D4		Over 70 with comorbidities			Hypoglycemia, liver injury, congestive heart failure			

*** Reasons for Gap**

Insufficient or Imprecise Information → A1=No studies, **A2**=Limited number of studies, **A3**=Sample sizes too small, **A4**=Estimate of effect is imprecise
Biased Information → B1=Inappropriate study design, **B2**=Major methodological limitations in studies
Inconsistency or Unknown Consistency → C1=Consistency unknown (only 1 study), **C2**=Inconsistent results across studies
Not the right information → D1=Results not applicable to population of interest, **D2**=Inadequate duration of interventions/comparisons, **D3**=Inadequate duration of follow-up, **D4**=Optimal/most important outcomes not addressed, **D5**=Results not applicable to setting of interest