

Methods Guide for Comparative Effectiveness Reviews

Prioritization and Selection of Harms for Inclusion in Systematic Reviews



This report is based on research conducted by the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers' Methods Workgroup. This research was funded through contracts from the Agency for Healthcare Research and Quality to the following Evidence-based Practice Centers: Pacific Northwest (290-2015-00009-I); Connecticut (290-2015-00012I); Duke (290-2015-00004-I); RAND (290-2015-00010-I); Johns Hopkins (290-2015-00006-I); Vanderbilt (290-2015-00003-I); Brown (290-2015-00002-I); Minnesota (290-2015-00008-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.

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

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policy makers, 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 is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance, contact epc@ahrq.hhs.gov.

Suggested citation: Chou R, Baker WL, Bañez L, Iyer S, Myers ER, Newberry S, Pincock L, Robinson KA, Sardenga L, Sathe N, Springs S, Wilt TJ. Prioritization and Selection of Harms for Inclusion in Systematic Reviews. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. AHRQ Pub No. 17(18)-EHC-034-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018. www.effectivehealthcare.ahrq.gov/reports/final.cfm
DOI: <https://doi.org/10.23970/AHRQEPCMETHODGUIDE1>.

Prepared by:

Scientific Resource Center
Portland, OR

Investigators:

Roger Chou, M.D., M.P.H.
William L. Baker, Pharm.D.
Lionel L. Bañez, M.D.
Suchitra Iyer, Ph.D.
Evan R. Myers, M.D., M.P.H.
Sydne Newberry, Ph. D.
Laura Pincock, Pharm.D., M.P.H.
Karen A. Robinson, Ph.D.
Lyndzie Sardenga, B.S.
Nila Sathe, M.A., M.L.I.S.
Stacey Springs, Ph.D., M.S.
Timothy J. Wilt, M.D., M.P.H.

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.

Strong methodological approaches to systematic review improve the transparency, consistency, and scientific rigor of these reports. Through a collaborative effort of the Effective Health Care (EHC) Program, the Agency for Healthcare Research and Quality (AHRQ), the EHC Program Scientific Resource Center, and the AHRQ Evidence-based Practice Centers have developed a Methods Guide for Comparative Effectiveness Reviews. This Guide presents issues key to the development of Systematic Reviews and describes recommended approaches for addressing difficult, frequently encountered methodological issues.

The Methods Guide for Comparative Effectiveness Reviews is a living document, and will be updated as further empiric evidence develops and our understanding of better methods improves.

If you have comments on this Methods Guide paper, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Gopal Khanna, M.B.A.
Director
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
Center for Evidence and Practice
Improvement
Agency for Healthcare Research and Quality

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

Lionel L. Bañez, M.D.
Task Order Officer
Evidence-based Practice Center Program
Center for Evidence and Practice
Improvement
Agency for Healthcare Research and Quality

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

Laura Pincock, Pharm.D., M.P.H.
Task Order Officer
Center for Evidence and Practice
Improvement
Agency for Healthcare Research and Quality

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers is as follows:

Hassan Murad, M.D., M.P.H.
Director, Mayo Clinic EPC
Rochester, MN

Jennifer Lin, M.D., M.C.R., FACP
Director, Kaiser Permanente EPC
Portland, OR

Karen Schoelles, M.D., SM, FACP
Director, ECRI-Penn EPC
Plymouth Meeting, PA

C. Michael White, Pharm.D., FCP, FCCP
Director, University of Connecticut EPC
Storrs, CT

Susan L. Norris, M.D., M.P.H., M.Sc.,
Investigator, Kaiser Permanente EPC
Portland, OR

Yoon Loke, M.D.,
Norwich Medical School, University of East
Anglia
Norwich, UK

Mary Butler, Ph.D., M.B.A.
Co-Director, Minnesota EPC
Minneapolis, MN

Lisa Hartling, Ph.D.
Director, University of Alberta EPC
Edmonton, Alberta, Canada

Meera Viswanathan, Ph.D.
Director, RTI-UNC EPC
Chapel Hill, NC

Mark Helfand, M.D., M.S., M.P.H.
Director, Scientific Resource Center
Portland, OR

Contents

Introduction	1
Methods	3
Approach.....	3
Literature Search and Review	3
Review of EPC Reports.....	3
Key Informant Interviews	3
Development of Recommendations	4
Results	5
Literature Search	5
Review of EPC Reports.....	5
Synthesis of KI Interviews	6
Use of Published Guidance for Prioritization and Selection of Harms.....	6
Criteria for Prioritizing Harms	6
Using Input From Stakeholders To Guide Prioritization and Selection of Harms.....	6
Using Literature and Other Data Sources To Guide Prioritization and Selection of Harms	6
Thresholds for Maximum Number of Harms To Be Reviewed.....	7
Inclusion of Composite Harms.....	7
Methods for Addressing Harms Not Included in the Original Review Protocol	7
Reporting of Methods for Prioritizing and Selection of Harms	8
Recommendations	9
Prioritization of Harms.....	9
Type of Harms To Include	10
Number of Harms To Include	11
Harms Not Specified in the Original Protocol	11
Reporting Methods Used To Select Harms.....	12
Discussion	13
References	14

Appendixes

- Appendix A. Guidance From Prior Chapter
- Appendix B. Key Informant Invite
- Appendix C. Key Informant Interview Guide
- Appendix D. Key Informant Interview Themes
- Appendix E. Key Informants
- Appendix F. Review of EPC Reports

Introduction

Guidance from within the Agency for Healthcare Research and Quality's (AHRQ) Evidence-based Practice Center (EPC) Program has long recognized the need for systematic reviews of interventions impacting health to provide balanced assessments that include evaluation of harms as well as benefits. However, synthesizing evidence on harms poses unique challenges: the assessment and reporting of harms is often suboptimal,^{1,2} the duration of studies is often too short to evaluate important long-term harms and have inadequate statistical power to evaluate serious but uncommon harms,^{3,4} patients enrolled in research studies are frequently at lower risk for harms than those encountered in clinical practice,⁵ and important data on harms may be unpublished or selectively reported.^{6,7}

In 2005, AHRQ funded a series of white papers on challenges in evidence synthesis that included an article on evaluation of harms.³ It highlighted unique challenges in finding and selecting data on harms, rating the quality of harms assessment and reporting, and synthesizing and displaying data from studies reporting harms. Subsequently, recommendations for synthesizing evidence on harms were further developed by a Methods Workgroup of EPC investigators convened by AHRQ; these recommendations were codified in 2010 as a chapter in the AHRQ EPC Program Methods Guide.⁸ Issues addressed by the Workgroup included the need to consider a broad range of data sources to evaluate harms, including observational studies as well as randomized controlled trials and unpublished as well as published data; the importance of using consistent and precise terminology on harms; the need to evaluate the quality of harms assessment and reporting distinctly from the rigor for assessing benefits; and challenges in synthesis, including evaluation of rare events, use of indirect comparisons, and pooling methods.

In 2015, AHRQ convened an EPC Methods Workgroup to update or expand upon prior guidance for assessing harms. Following initial deliberations, the Workgroup elected to focus on updating guidance on prioritization and selection of harms to assess in systematic reviews. The Workgroup determined that guidance from the prior harms chapter (Appendix A) remains current. However, although the 2010 harms chapter recommended that EPC systematic reviews "always assess harms that are important to clinicians and patients," the Workgroup found that it lacked more specific recommendations regarding how to select the harms to be included in an EPC review, and noted that selection and prioritization of harms in EPC reviews poses important challenges. EPC reviews frequently address many interventions, which could result in many potential harms (e.g., dozens) to review. Unlike benefits, which are often similar across interventions used to treat a given condition (e.g., medications, non-pharmacological therapies, and surgery for low back pain are all aimed at improving pain and function), different interventions given for the same condition are frequently associated with a large number of diverse harms. For example, medications for low back pain are typically associated with a set of harms distinct from those associated with surgery, and different medications are each associated with unique harms. Other issues include whether to assess composite harms (e.g., "serious harms" or "withdrawal due to adverse events"), which might facilitate comparisons between interventions with dissimilar harms, and how to address harms that are not specified in the original protocol but encountered during the review process. Workgroup members noted that including all potential harms in these cases is not feasible, and can make it difficult for users of EPC reviews to reach bottom-line conclusions regarding harms or determine the balance of benefits to harms. Workgroup members noted that clearer methods would be helpful for supporting the decisions made regarding selection of harms and help focus EPC reviews on the outcomes of greatest importance, potentially increasing their usability.

The purpose of this report is to provide guidance on prioritization and selection of harms for inclusion in systematic reviews. The immediate intended audience of this guidance is the EPC program, though we hope it may be useful to all systematic reviewers and those who commission or use systematic reviews

Methods

Approach

We assembled a workgroup of 12 methodologists from AHRQ, the EPC program, and the Scientific Resource Center (SRC) to develop recommendations on selection and prioritization of harms, building on work by a prior EPC Workgroup. Members participated in twice monthly teleconference calls over the span of 11 months and sought information on selection and prioritization of harms through a literature search for empiric research and published guidance, a review of EPC reports to understand how harms have been selected in the past, and interviews with experts in the conduct of systematic reviews or users of systematic reviews (Key Informants) on experiences and suggestions for selection and prioritization of harms, in order to inform the development of consensus recommendations.

Literature Search and Review

The SRC provides support for the AHRQ EPC Program for the advancement of scientific methods, strategic planning, peer review, topic nomination and education. As part of this work, the SRC curates a bibliographic database of nearly 10,000 citations on the methodology of systematic reviews and comparative effectiveness research, dating back to the 1950s.⁹ On November 10, 2015, the SRC librarian performed a keyword search and a descriptor search for “Harms/Adverse Events” in the SRC Methods Library database (n=357). The citations were filtered from a publication date of 2007 on, so as to only include more recent articles, including studies published since the prior harms chapter in the AHRQ EPC Program Methods Guide (n=257).⁸ Two members (RC, NS) of the workgroup then conducted a dual review of the citations, seeking articles that could provide guidance on the methods for selecting and prioritizing harms for inclusion in systematic reviews, or that reported empiric research in that area. Because we were seeking literature that could inform discussions and context and anticipated that empiric research would be sparse, we did not apply strict eligibility criteria.

Review of EPC Reports

A workgroup member from the SRC (LS) reviewed 18 EPC reports to determine year of publication, key questions related to harms and the harms that were assessed, methods used to prioritize or select harms, the data sources used to identify evidence on harms, and main findings regarding rates of harms. We categorized the harms assessed as “individual” or “composite harms;” composite harms included measures such as any harm, serious harms, withdrawal or discontinuation due to adverse events, or similar. We selected EPC reports published since the year 2014 to ensure the sample represented recent methods. We did not review a random sample of EPC reports, but instead selected a sample of reports from various EPCs with a diversity of types of interventions evaluated and compared. Information from the EPC reports was abstracted into an Excel spreadsheet, which was provided to Workgroup members to inform discussions.

Key Informant Interviews

The SRC compiled a list of methodologist key informants (KIs) to interview who had experience in conducting, commissioning, or using systematic reviews. On February 19, 2016, the SRC sent email invites to (n=14) priority KIs with expertise in conduct of reviews,

assessment of harms in reviews, and/or representing organizations that commission and use systematic reviews, providing them with background on the project and the purpose of the requested interview (Appendix B).

The workgroup lead (RC), with input from the group, compiled a list of 12 targeted questions with the aim of informing our discussion on the selection and prioritization of harms in systematic reviews (Appendix C). The interview guide was sent to KIs prior to their interview and covered these general topics:

- Use of published guidance for prioritization and selection of harms
- Criteria for prioritizing harms
- Use of input from stakeholders to guide prioritization and selection of harms
- Use of literature and other data sources to guide prioritization and selection of harms
- Thresholds for maximum number of harms to be reviewed
- Inclusion of composite harms
- Methods for addressing harms not included in the original protocol
- Reporting of methods for prioritizing and selection of harms.

KIs also completed a conflict of interest form prior to their participation; none were determined to have conflicts that precluded their participation. Over the span of 5 months (March 2016 – July 2016), workgroup members conducted 5 telephone interviews with 6 KIs, lasting 60 minutes each. Each interview was recorded, transcribed, and sent to its respective KI as an opportunity for further elaboration, clarification, and corrections. The notes were transcribed and one Workgroup member identified common themes in the responses to interview questions across the KIs, as well as areas in which responses differed. A second Workgroup member reviewed the notes and common themes for clarifications and additions. Notes from each interview and a document summarizing the themes were then presented to the Workgroup for further discussion (Appendix D).

Development of Recommendations

All Workgroup members reviewed the results of the literature search, review of EPC reports, and notes from the KI interviews and summarized themes, which were discussed on regularly scheduled conference calls. A draft set of recommendations on selection and prioritization of harms in systematic reviews was developed by the workgroup lead (RC) and distributed to the Workgroup for further discussion and feedback. Given the lack of strong empiric evidence to guide recommendations in this area, the Workgroup sought to reach consensus on all recommendations.

Results

Literature Search

After screening 257 citations from the SRC methods research database and reviewing 108 full-text articles, we identified no empiric research on the utility or validity of different methods to prioritize harms in systemic reviews to inform our discussions. Although several articles provided general guidance on assessment of harms or on prioritization of outcomes for systematic reviews, none provided recommendations specifically on selection and prioritization of harms, other than the prior EPC methods work.⁸ The harms chapter of the EPC Methods Guide recommended that reviewers assess the harms that are important to decision makers and users of the intervention under consideration; it noted that high-priority harms are the most serious adverse events and may include common adverse events or other adverse events important to clinicians and patients. It suggested that systematic review authors use prior reviews, safety reports from the US Food and Drug Administration, post-marketing surveillance databases, and input from technical experts and patients to identify and prioritize harms to be evaluated.

We also reviewed articles providing general guidance on conduct of systematic reviews and synthesis of evidence, but found little guidance on selection and prioritization of harms. For example, regarding selection of harms, the Cochrane Handbook notes the harms selected for a review depend on the study question and the therapeutic or preventive context, and that the reviewer could opt for a narrow focus (e.g., one or two known or a few of the most serious adverse effects that are of special concern to patients and health professionals) or a broad focus (e.g., the 5 to 10 most frequent adverse effects, all adverse effects that either the patient or clinician consider to be serious, or organized by category [e.g., diagnosed by lab results or patient-reported symptoms]).¹⁰ The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group recommends that guideline developers prioritize outcomes (both beneficial and harmful), which can be done through solicitation of panel member and stakeholder input and using a 1-9 numerical rating system.¹¹ The outcomes rated highest priority are the ones that the guideline development group will focus on in assessing the balance of benefits to harms and informing recommendations. The GRADE working group recommends that summary of findings tables focus on no more than 7 of the most patient-important outcomes (including both beneficial and harmful outcomes), in order to avoid overwhelming the reader, while providing information on the most critical outcomes.¹²

Review of EPC Reports

We reviewed 18 EPC reports that addressed a range of intervention types (e.g., medical, surgical, diagnostic testing, informatics, behavioral therapy) and conditions (e.g., cancer, musculoskeletal, surgical, psychological, lipids, obstetric, neurological, otolaryngologic). All of the reports had key questions related to harms. No EPC report described the method used to select or prioritize harms, though most reported on serious and common harms, implying that severity and frequency guided decisions regarding which harms to include. Few EPC reports described results for composite harms such as “any adverse event,” “withdrawal due to adverse event,” or “any serious adverse event”; rather, the reports generally focused on specific adverse events, sometimes categorizing their severity. No EPC report described using a formal prioritization process or the sources used to inform decisions regarding which harms to include.

Synthesis of KI Interviews

Use of Published Guidance for Prioritization and Selection of Harms

None of the KIs reported using published guidance for prioritization and selection of harms. Although the KIs were generally aware of GRADE methods for prioritization of outcomes, none reported using GRADE methods to prioritize harms, and few had experience applying a formal GRADE prioritization process to selection of outcomes in general. Several KIs were aware of published guidance from the Cochrane Collaboration and AHRQ on assessing harms in systematic reviews, but were not aware of specific guidance on prioritization and selection of harms.

Criteria for Prioritizing Harms

All the KIs noted that systematic reviews should prioritize harms that are of greatest importance to decision makers. They noted that these typically include serious harms as well as less-serious but common harms. The KIs noted that severity of harms is often poorly or inconsistently defined, which makes determination of whether a harm is “serious” a challenge, though the KIs also noted that there are published definitions for categorizing seriousness of harms (e.g., the Food and Drug Administration criteria for reportable “serious” adverse events).¹³ The KIs noted that the quality or quantity of evidence should not be an important factor in selection and prioritization of harms; rather, they emphasized the need to select and prioritize harms that are important to decision makers, regardless of the evidence and evidence sources available.

Using Input From Stakeholders To Guide Prioritization and Selection of Harms

All KIs described obtaining input from clinical and content experts to inform decisions regarding prioritization and selection of harms. Although most KIs described a relatively informal process (e.g., soliciting general feedback from a panel of stakeholders on outcomes to be addressed in a conference call or electronically), others described a more formal process in which KIs were asked to rank or rate outcomes related to harms. Several KIs also described using input from patient stakeholders. In some cases, patients participated in a larger group with clinical or content experts and in others, patients provided input separately. The KIs noted that a challenge in engaging stakeholders to prioritize harms is that clinicians and patients could prioritize harms differently.¹⁴ One KI described a process in which patients were asked to rank/prioritize outcomes and described challenges in interpreting or using the findings, such as patients rating all outcomes as similarly high priority (e.g., mortality and an intermediate laboratory outcome both prioritized similarly) or patients having difficulty understanding the systematic review process or the scientific issues. Another KI noted that her organization had convened a group of patient stakeholders who are to receive training in systematic review methods and would be asked to provide input for multiple reviews on an ongoing basis.

Using Literature and Other Data Sources To Guide Prioritization and Selection of Harms

All of the KIs noted that a broad range of data sources should be utilized to inform decisions regarding which harms to consider for inclusion. Suggested data sources included randomized controlled trials, observational studies (including pharmacoepidemiological studies performed on large databases), and information from regulatory agencies and other groups that collect

postmarketing information and case reports on adverse events (e.g., the FDA’s MedWatch program).¹⁵ The KIs noted that reviewers should not rely solely on randomized trials since they are often underpowered to detect uncommon harms, frequently too short to evaluate long-term harms, and often enroll “ideal” populations at low risk for harms. The KIs noted that case reports may identify potentially serious harms that are very uncommon; however, they also noted that it can be difficult to determine causality from such studies.

Thresholds for Maximum Number of Harms To Be Reviewed

The KIs were generally aware that the GRADE Working Group has suggested a maximum number of outcomes to include when summarizing the evidence. However, they felt that it was difficult to apply a maximum threshold for harms to be included in EPC reports, given the large number of interventions and comparisons that are often being evaluated. In addition, the KIs noted that even for a single comparison, limiting to a maximum of 7 beneficial and harmful outcomes as recommended by the GRADE Working Group could result in only 3 or 4 harms, which they felt was fewer than necessary to adequately assess the harms of most interventions. Nonetheless, the KIs agreed that it is important to focus the EPC reports on the most critical harms, without applying a specific threshold for the maximum number to be evaluated, to help make the reports more usable.

Inclusion of Composite Harms

The KIs generally felt that inclusion of composite harms could be helpful in facilitating comparisons between interventions with dissimilar harms. They noted that composite harms, like other composite outcomes, consist of a variety of different harms, and must be interpreted with caution. They suggested that if composite harms are included, it would generally be more useful to focus on more severe harms, as indicated by “any serious harm” or “withdrawals due to adverse events,” rather than composite harms that include less serious events (e.g., “any adverse event”), which may be more difficult to interpret. The KIs also noted that composite harms should be interpreted in conjunction with data on individual harms.

Methods for Addressing Harms Not Included in the Original Review Protocol

The KIs noted that during the review process, reviewers may encounter or become aware of potentially relevant harms not considered in the prioritization process and included in the review protocol. This discovery could be due to the publication of new data or analyses, or patterns/data that the reviewers observe in the course of conducting the review. The KIs suggested that reviewers should be open to including information on harms not identified during the protocol development phase, and be prepared to modify the study protocol to note their inclusion. However, they also expressed the belief that data and analyses regarding such harms could generally be considered hypothesis generating and presented as such when appropriate. KIs indicated that for harms not included in the review protocol for which data appear compelling, reviewers should consider proposing future research to clarify potential associations. The reviewers noted examples in which such harms ended up not being clearly confirmed in prospective studies (e.g., increased myocardial infarction with thiazolidinediones)^{16,17} as well as examples in which such harms have been confirmed in subsequent analyses (e.g., increased myocardial infarction risk with cyclo-oxygenase-2-selective non-steroidal anti-inflammatory agents).¹⁸

Reporting of Methods for Prioritizing and Selection of Harms

The KIs agreed that systematic reviews typically do not report methods used to select or prioritize harms; this was consistent with our review of EPC reports. The KIs felt that it would be helpful for systematic reviews to report any prioritization methods used, including how stakeholders were engaged, criteria used to determine which harms were included (e.g., seriousness, frequency), and the criteria used to select included harms. The KIs noted that providing this methodologic information would help readers better understand the basis for prioritization decisions.

Recommendations

Prioritization of Harms

1. Include harms that are of greatest importance to decision-makers.

The workgroup recommends that EPC reviews include the harms judged to be of greatest importance to decision-makers, including patients, clinicians, policymakers, and other stakeholders. Typically, these will be serious harms as well as less serious but common and/or bothersome harms (e.g., angiotensin converting enzyme inhibitor-induced cough). Using the FDA definition, serious harms are those that result in death, are life-threatening, result in hospitalization or prolongation of an existing hospitalization, result in persistent or significant incapacity or ability to perform normal life functions, or are congenital anomalies or birth defects.¹⁹ Other harms may also be considered serious when judged to jeopardize the patient or study participant and may require medical or surgical intervention to prevent one of the outcomes listed above. From the perspective of a decision-making framework, a harm may be considered “important” if the probability of that harm occurring compared to expected benefits would impact recommendations about the use of the intervention. An exception to routinely including all harms of greatest importance is reviews that focus on a specific, pre-defined harm or harms (this is not typical for EPC reviews); in these cases, the scope of the review should be clearly explained.

2. Use a prioritization process to help narrow the number of harms included in a review.

Recognizing that it will often not be feasible to include all potential harms in an EPC review, the workgroup recommends that EPCs utilize a process to prioritize the harms of greatest importance to be reviewed. Generally speaking, the harms prioritized in this process will be included in summary of evidence tables, along with prioritized benefits. Using a prioritization process will help strengthen the rationale for the harms that are selected for review and provide a basis for the selection decisions that are made.

3. The specific prioritization process used can vary. The prioritization may be informal (e.g., input or informal interviews with experts in the field, patients, and other stakeholders, literature search, review of information from regulatory agencies) or more formal (e.g., Delphi or GRADE-like scoring process).

Although the workgroup suggests that EPC utilize a prioritization process, it found insufficient evidence to recommend a specific prioritization method. The workgroup suggests that EPCs obtain input from stakeholders, including clinical/technical experts, policymakers, and patients; perform a literature search; and review information from regulatory agencies to inform the prioritization process. Although more formal prioritization methods may be useful (e.g., formal consensus process or use of a GRADE-like scoring/prioritization method), the workgroup concluded that it is unclear whether using such methods results in more appropriately selected/prioritized harms than less formal processes. Whenever possible, the EPC should obtain input from patients, to ensure that outcomes reflect their priorities.²⁰ As noted by some KIs, incorporating stakeholder input in a more formal process also could present a challenge, for example when different stakeholders prioritize harms differently or when very serious clinical outcomes and minor harms (e.g., laboratory based intermediate outcomes) are prioritized

similarly. In addition, utilizing such processes impact the time and resources required to conduct the review. Therefore, until more data are available on the effects of using more formal prioritization processes on the usefulness and credibility of systematic reviews, the workgroup concluded that a recommendation for their routine use was not warranted.

4. The method used to prioritize harms should be concordant with methods used to select outcomes related to benefit.

The workgroup recommends that the methods used to prioritize and select harms be concordant with the methods used to prioritize and select beneficial outcomes, given that the principles underlying the prioritization of outcomes are similar, whether they are to measure beneficial or harmful effects. The workgroup acknowledged that prioritization of beneficial outcomes is often more straightforward than for harmful outcomes since the expected benefits for different interventions administered for the same condition are often similar and the potential beneficial effects of interventions are often well-understood. However, as for harmful outcomes, there may be many potential beneficial outcomes to consider. It may be difficult to distinguish harms from failed treatments (e.g., myocardial infarction in patients on statin therapy); whether an event is classified as a benefit or harm may depend on the intended effect of the treatment and the perspective of the decision maker.

Types of Harms to Include

5. Routinely include serious harms or less serious but frequent or bothersome harms, or describe why they are not included.

As noted above, the workgroup recommends that EPC reviews routinely include serious harms or less serious but frequent or bothersome harms. In some cases, EPC reviews may not include all such harms. This could be because the harms are well-established and do not require another review; the intervention is not thought to be associated with major harms (e.g., eyeglasses for decreased visual acuity, hearing aids for hearing loss, ultrasound for musculoskeletal conditions); or the review is focused on a particular harm or harms. When applicable, such circumstances should be explained. In general, intermediate outcomes (e.g., changes in laboratory values or physiological parameters) are considered lower priority than patient-centered health outcomes (e.g., mortality or outcomes related to morbidity, quality of life, or function). EPC reviews may consider inclusion of intermediate outcomes related to harms when data on associated clinical outcomes are sparse and the association between intermediate outcomes and clinical harms is well established (e.g., severe anemia or neutropenia).

6. Composite adverse events may help facilitate comparisons across interventions; routinely consider including “serious adverse events” or “withdrawal due to adverse events,” particularly when evaluating head-to-head comparisons.

The workgroup recommends that EPC reviewers consider including composite adverse events, which may help facilitate head-to-head comparisons, particularly for interventions associated with dissimilar harms. The workgroup suggests that EPC reviews focus on indicators of more severe harms (e.g., “serious” adverse events or “withdrawal due to adverse events”), given the composite nature of these outcomes, as it is more difficult to interpret the clinical meaningfulness of less severe harms. The workgroup recommends that EPC reviews not focus solely on composite adverse events; rather, composite harms should be interpreted in conjunction

with data on the individual harms that comprise these composite outcomes. EPC reviews should record the definitions used for composite harms, which often vary across studies.

7. For reviews that involve effects of diagnostic tests, consider inclusion of over-diagnosis and overtreatment, as well as other harms related to diagnostic testing (e.g., false positives and false negatives and effects thereof, labeling, and others).

Although intermediate outcomes are generally considered lower priority than clinical outcomes, for certain interventions (e.g., those addressing diagnostic tests), the workgroup suggests that EPC reviews consider inclusion of intermediate measures of harm such as over-diagnosis or overtreatment as a result of testing. Such outcomes may help identify important negative downstream effects of testing that are otherwise difficult to capture. The workgroup acknowledges challenges in measuring these outcomes, and variability in the methods used.²¹⁻²³ Other harms associated with diagnostic tests include false-positives and false-negatives and the consequences of such findings, labeling, and others.

Number of Harms To Include

8. A reasonable rule of thumb is to limit to 5-10 prioritized harms for each comparison involving two interventions, though there is no preset threshold for the number of harms selected for a review.

Given the large number of interventions and comparisons that may be included in an EPC review, the number of potential harms to be reviewed may make it difficult for users to process and interpret the findings. Therefore, the workgroup recommends that EPC reviews limit the number of prioritized harms to be reviewed. The workgroup felt that using the suggested GRADE maximum threshold of 7 beneficial and harmful outcomes would frequently result in exclusion of potential important harms. Instead, it suggests that EPC reviewers utilize an approach that is based on the number of comparisons. For each comparison involving two interventions, the workgroup suggests that the EPC aim for 5-10 or fewer prioritized harms (including individual as well as composite harms), though for some comparisons it may be appropriate to prioritize more than 10 harms. Across each comparison, to the extent possible the workgroup suggests that EPCs identify common prioritized harms, in order to limit the total number of harms to be assessed. For reviews in which there are many comparisons and potential harms, the workgroup suggests that reviewers aim for a number of harms selected for each comparison involving two interventions on the lower end of the range.

Harms Not Specified in the Original Protocol

9. Be prepared to add harms to the review that were not specified in the original protocol or identified in the prioritization process; in some cases, findings for such harms will be considered hypothesis generating.

The workgroup recommends that EPC reviews be prepared to incorporate harms not in the original protocol into the review. Such harms may be identified during the course of data analysis of included studies, or via outside sources (e.g., new published study, regulatory agency action). Because these harms are not pre-specified, their addition should be recorded as a protocol modification. In addition, the workgroup suggests that EPC reviews clearly indicate findings related to harms that are not specified in the original protocol. EPCs should interpret findings related to such harms in the context of other information, including the plausibility of

biological mechanisms of action, pharmacokinetic and pharmacodynamic data, the magnitude of effect, the precision of estimates, the statistical significance of findings, and other data on the harm that may have previously been overlooked or unidentified.²⁴ In some cases (e.g., isolated case reports, small magnitude of effects, imprecise estimates, high confounding potential, no biologically plausible mechanism), findings for such harms will be considered hypothesis-generating.

Reporting Methods Used to Select Harms

10. Report the methods used to prioritize harms, differentiate serious from frequent but less serious harms, and indicate interventions for which serious harms are not believed to be an issue and why.

The workgroup recommends that EPC reviews describe methods used to prioritize harms, including the composition of stakeholder groups providing input, literature search methods, and other data sources. In addition, EPC reviews should describe the prioritization process, whether informal or more formal. EPC reviews should differentiate which harms are considered serious and those considered less serious but of high frequency or most bothersome. In situations in which serious harms are not included, EPC reports should provide the reason (e.g., the intervention is not believed to be associated with serious harms, serious harms have already been established, the review is scoped to focus on a specific harm or harms).

Discussion

Selection or prioritization of harms in EPC reviews is an important challenge that has not previously been addressed in depth in the EPC Methods Guide. Although EPC reviews seek to be comprehensive and provide balanced assessments of benefits and harms, inclusion of multiple interventions and comparisons often results in consideration of many potential harms, which could be overwhelming to users. A review of EPC reports indicate that they provide little or no information regarding how harms were selected. A search of the literature found little guidance on selection and prioritization of harms in systematic reviews. A limitation of this article is the lack of empiric research on prioritization to guide development of recommendations, a small sample of KIs providing input, and relatively narrow literature search strategy. This article provides guidance developed by a workgroup of EPC methodologists on selection and prioritization of harms. Key recommendations include: routinely focusing on serious as well as less serious but frequent or bothersome harms; routinely engaging stakeholders and using literature searches and other data sources to identify harms of importance; using a prioritization process (whether formal or less formal) to inform selection decisions; and describing the methods used to select and prioritize harms. The workgroup identified methods for assessing the quality of harms reporting and determining which types of studies to use to evaluate harms as high priority topics for future guidance development. The workgroup recognizes that the data supporting the recommendations in this article are sparse and that follow-up to assess the impact of the recommendations on reporting, usefulness/usability of reports, and appropriateness of prioritization decisions is needed.

References

1. Haidich AB, Birtsoy C, Dardavessis T, et al. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. *J Clin Epidemiol.* 2011;64(2):124-35. doi: 10.1016/j.jclinepi.2010.03.005.
2. Hodkinson A, Kirkham JJ, Tudur-Smith C, et al. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ Open.* 2013;3(9):e003436-2013-. doi:10.1136/bmjopen-2013-003436.
3. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med.* 2005;142(12 Pt 2):1090-9. doi: 10.7326/0003-4819-142-12_Part_2-200506211-00009.
4. Bhaumik DK, Amatya A, Normand SL, et al. Meta-Analysis of Rare Binary Adverse Event Data. *J Am Stat Assoc.* 2012;107(498):555-67. doi:10.1080/01621459.2012.664484.
5. Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet.* 2005;365(9453):82-93. doi:10.1016/S0140-6736(04)17670-8.
6. Hodkinson A, Gamble C, Smith CT. Reporting of harms outcomes: a comparison of journal publications with unpublished clinical study reports of orlistat trials. *Trials.* 2016;17(207). doi:10.1186/s13063-016-1327-z.
7. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ.* 2014;349:g6501. doi:10.1136/bmj.g6501.
8. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol.* 2010;63(5):502-12. doi: 10.1016/j.jclinepi.2008.06.007.
9. Agency for Healthcare Research and Quality. AHRQ Secure Site Methods Article Alert - EHC Scientific Resource Center EPC-SRC.org. 2016. <http://epc-src.org/methodsLibrary/artAlert.cfm> (accessed October 2, 2017)
10. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* The Cochrane Collaboration, 2011. <http://handbook.cochrane.org> (accessed October 2, 2017).
11. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011;64(4):395-400. doi:10.1016/j.jclinepi.2010.09.012.
12. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes. *J Clin Epidemiol.* 2013;66(2):158-72. doi:10.1016/j.jclinepi.2012.01.012.
13. U.S. Food and Drug Administration. CFR - Code of Federal Regulations Title 21 Sec. 312.32; 2016. <https://www.accessdata.fda.gov/script/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=312.322016> (accessed October 2, 2017)
14. Montgomery A, Fahey T. How do patients' treatment preferences compare with those of clinicians? *Qual Health Care.* 2001;10(1). doi:10.1136/qhc.0100039.
15. Craigle V. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. *J Med Libr Assoc.* 2007;95(2):224-5. doi:10.3163/1536-5050.95.2.224.

16. Dormandy JA, Charbonnel B, Eckland DJ, Erdmann E, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAZone Clinical Trial In macroVascular Events): a randomised controlled trial. *Lancet*. 2005;366(9493):1279-89. doi:10.1016/S0140-6736(05)67528-9.
17. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet*. 2009;373(9681):2125-35. doi:10.1016/S0140-6736(09)60953-3.
18. Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ*. 2011;342(7086). doi:10.1136/bmj.c7086
19. Administration USFaD. What is a Serious Adverse Event? 2016. <http://www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm> (accessed October 2, 2017)
20. Garces JPD, Lopez GJP, Wang Z, et al. Eliciting patient perspective in patient-centered outcomes research: a meta narrative systematic review. A report prepared for the Patient-Centered Outcomes Research Institute. Rochester: Mayo Clinic. 2012. <https://www.pcori.org/assets/Eliciting-Patient-Perspective-in-Patient-Centered-Outcomes-Research-A-Meta-Narrative-Systematic-Review.pdf> (accessed October 2, 2017)
21. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol*. 2014;65(6):1046-55. doi:10.1016/j.eururo.2013.12.062.
22. Gibbons RD, Amatya A. *Statistical Methods for Drug Safety*. Chapman and Hall/CRC: Boca Raton, FL; 2015.
23. Jha S. An Overview of Methodologies in Detecting Overdiagnosis. *Acad Radiol*. 2015;22(8):1046-7. doi:10.1016/j.jacr.2015.06.008.
24. Avorn J. Evaluating drug effects in the post-Vioxx world there must be a better way. *Circulation*. 2006;113(18):2173-6. doi:10.1161/CIRCULATIONAHA.106.625749.

Appendix A. Guidance From Prior Chapter¹

Summary of key points on assessment of harms in Comparative Effectiveness Reviews

- Assess all important harms, whenever possible.
- Use multiple sources of information, including clinical experts and stakeholders, to identify important harms.
- Use consistent and precise terminology when reporting data on harms, and avoid terms implying causality unless causality is reasonably certain.
- Gather evidence on harms from a broad range of sources, including observational studies, particularly when clinical trials are lacking; when generalizability is uncertain; or when investigating rare, long-term, or unexpected harms.
- Do not assume studies adequately assess harms because methods used to assess and report benefits are appropriate; rather, evaluate how well studies identify and analyze harms.
- Be cautious about drawing conclusions on harms when events are rare and estimates of risk are imprecise.
- Include placebo-controlled trials, particularly for assessing uncommon or rare harms, but be cautious about relying on indirect comparisons to judge comparative risks, and evaluate whether studies being considered for indirect comparisons meet assumptions for consistency of treatment effects.
- Avoid inappropriate combining of data on harms, and thoroughly investigate inconsistent results.

1. Chou R, Aronson N, Atkins D, et al. AHRQ Series Paper 4: Assessing Harms When Comparing Medical Interventions: AHRQ and the Effective Health-Care Program. *Journal of clinical epidemiology*. 2010;63(5):502-12.

Appendix B. Key Informant Invite

Dear Dr. [insert name],

We are conducting research on the various approaches used to select and prioritize harms to be included in systematic reviews. This is an extension of a Guidance chapter we produced in 2008 that examined methods and guidance for Harms reporting. This was identified as an important area given the potentially large number of harms that could be assessed in many reviews. As part of this Agency for Healthcare Research and Quality (AHRQ) funded project, we are having discussions with thought leaders in the field who conduct, commission, or use systematic reviews to can help inform guidance on prioritization of harms for systematic reviews conducted by the AHRQ Evidence-based Practice Centers (EPC).

Because of your experience in conducting, commissioning, or using systematic reviews we would like to schedule a time to speak with you. Please let us know if there is a different person in your organization that you think we should contact instead.

Your participation would involve a 60-90-minute panel interview. In this discussion we hope to learn more about the approaches taken throughout your program to select and prioritize harms in systematic reviews. *To help frame this discussion please feel free to read the 2008 EPC methods guidance on Harms, which contains limited guidance on this topic here: [Assessing Harms when Comparing Medical Interventions](#)*

If you are able to participate, please respond to our doodle poll with your availability at [Doodle Link](#)

If you are unable to make any of the above times, please let us know and we may be able to arrange another meeting time.

Please confirm if whether or not you able to participate in this project by Friday March 4, 2016. If you are able to participate, please complete and submit an AHRQ agreement to participate form and conflict of interest disclosure form here: [Secure Site Link](#)

Thank you for your consideration. If you have any questions, or would like additional information, please contact Lyndzie Sardenga at methods@epc-src.org or 503.220.8262 x58609.

Sincerely,

Roger Chou, M.D.
Director
Pacific Northwest Evidence-based Practice Center
AHRQ Effective Health Care Program

Sent on behalf of Roger Chou by the AHRQ Scientific Resource Center

Appendix C. Key Informant Interview Guide

Introduction

When conducting systematic review of medical interventions, it is important to evaluate harms as well as benefits in order to have balanced assessments. However, selecting harms for evaluation in systematic reviews can be a challenge, given the potentially large number of harms, particularly when comparing multiple interventions or when comparing different types of interventions (e.g., surgical vs. pharmacological). In addition, there are non-specific harms (e.g., any adverse event, serious adverse events) and intervention-specific harms, as well as differences in the evidence available to assess different harms. AHRQ has convened a methods workgroup that seeks to provide guidance on prioritization and selection of harms to EPC's conducting systematic reviews, given the relative lack of guidance in this area.

The goal of this discussion is to examine methods, strategies, and principles for prioritizing and selecting harms to be included in systematic reviews. We would like to draw on your experiences in conducting and using systematic reviews, as well as general methodological expertise. Some questions we will use to guide the interview are listed below. However, depending on your input and direction of the interview we may ask additional questions and may not address every question on the interview (you can provide additional comments after the call via email for questions not addressed on the call).

- 1) Do you use or refer to published guidance on how to prioritize and select harms for review? If so, what is the source of the guidance and what are the main principals?
- 2) Outside of published guidance, are there other principles that you think are useful for guiding how to prioritize and select harms for review?
- 3) Do you think it is reasonable to apply a maximum threshold in terms of the number of harms that are included in review? If so, what would you consider a reasonable maximum threshold? Does the number of beneficial outcomes reviewed impact the number of harmful outcomes reviewed?
- 4) Are there general (non-specific) harms (e.g., any adverse event, any serious adverse event, withdrawal due to adverse event) that you think should routinely be included as an assessed harm?
- 5) For specific harms, what criteria do you think are useful for prioritizing which harms to review? E.g., seriousness (how is "seriousness" determined), frequency, other?
- 6) Do you use GRADE or another formal process to prioritize which harms to review. If so, what has been your experience using such methods?
- 7) Does the type of evidence that is available for particular harms influence what types of harms should be selected for review? Do you prioritize data available from certain types of study designs (e.g., RCTs, cohort studies) or from certain types of sources (e.g., published literature vs. grey literature sources e.g. FDA or other regulatory agency). If so, how?

- 8) Sometimes harms for an intervention may not be known before an analysis has been performed (e.g., MI with rofecoxib). Should reviewers routinely try to incorporate important unanticipated harms in their review?
- 9) How should input from technical and clinical experts and patients or patient advocacy groups be used in selecting and prioritizing harms for review?
- 10) What other sources are useful for determining which harms should be included in a review?
- 11) Are there situations where the selection and prioritization of harms may change during the course of a review? What types of situations would it be reasonable to select different or additional harms for review?
- 12) Do you think systematic reviews adequately explain how harms have been selected for review? What types of information should be provided for users/readers of the review with regard to selection and prioritization of harms?

Appendix D. Key Informant Interview Themes

1. Do you use or refer to published guidance on how to prioritize and select harms for review? If so, what is the source of the guidance and what are the main principals?
 - *Only uses GRADE: which doesn't focus on harms specifically, but prioritizing outcomes in general.*
 - *Need to emphasize to groups that are prioritizing that they should include harms.*
 - *Have not found appropriate guidance on prioritization and selection of harms to use*
 - *Has turned to engaging decision makers to see what harms are most important for decisions being made. What outcomes will impact the decision? These decision makers are usually clinical experts.*
 - *Not currently but will start using GRADE to rate harms and benefits.*

2. Outside of published guidance, are there other principles that you think are useful for guiding how to prioritize and select harms for review?
 - *Most of the time targeting specific harms when conducting post marketing requirements, and already know which harms you want to look at. Size of the effect, source of the data, populations.*
 - *Focus on persons affected by recommendations/ what matters to them*
 - *2 big categories:*
 - *Anticipated harms: (specific vs non-specific)*
 - *Unanticipated harms: difficult to say how you would measure unanticipated harms because you can anticipate them.*
 - *Clinical judgment is important in thinking about harms in their frequency and how serious the event is, that is what our committee has come up with.*

3. Do you think it is reasonable to apply a maximum threshold in terms of the number of harms that are included in review? If so, what would you consider a reasonable maximum threshold? Does the number of beneficial outcomes reviewed impact the number of harmful outcomes reviewed?
 - *Focus on two or three of the most important safety outcomes. But if the purpose of a review is to update the labeling, then there shouldn't be a limit in terms of how many safety outcomes you want to put on the label.*
 - *No limit imposed*
 - *Groups do bring common / high burden AE to the table, perhaps this is more ad hoc than ideal*
 - *Not sure this is an appropriate approach. Harms are harms. Focus is to help patients make decisions, present findings to help them make decisions.*
 - *Fixed numbers are very difficult, should evaluate 2-3 harms at most.*
 - *Uses a three-pronged approach to do this: Consults with clinical experts about which harms they're most concerned about. Hopefully there are just one or two.*
 - *Patient's voice and what they are most concerned about needs to be considered as well.*
 - *Stick to a number below 7 for all outcomes.*
 - *Never encountered a situation where there are so many harms*

4. Are there general (non-specific) harms (e.g., any adverse event, any serious adverse event, withdrawal due to adverse event) that you think should routinely be included as an assessed harm?

- *Sometimes consider composite safety AE. One considered frequently is major adverse cardiovascular events. Combine non-fatal MI, non-fatal stroke, congestive heart failures, etc. in clinical trial settings and visual study settings.*
 - *Hesitant to combine serious AE in a composite way.*
 - *Comparing harms across different interventions might not be scientifically justified.*
 - *Maybe for non-specific harms, with serious adverse events might give some composite information; but be careful with interpretation of composite outcomes, can be challenging.*
 - *Collect some information on serious adverse events and total number of events. There is a weakness in composite measures. Can split list into serious AE or collection of adverse events 5 most common, serious ones are rare and frequent ones are mild or not life threatening.*
 - *Composite outcomes present problems. Number of patients who stop treatment because of side effects is a homogeneous measure, otherwise, lumping things together is risky.*
 - *They are helpful, it does give you an idea looking at drugs and AE, one thing to be sure of is, when you have this catch-all category make sure it's well defined (e.g. what makes an adverse event 'serious')*
5. For specific harms, what criteria do you think are useful for prioritizing which harms to review? E.g., seriousness (how is "seriousness" determined), frequency, other?
- *Depending on the amount of resources available, must prioritize based on severity, strength of evidence suggesting harm, type and quality of supporting evidence, and what the level of association is.*
 - *Prioritization works well; as seen with benefits.*
 - *Prioritization exercises to identify top outcomes*
 - *Need to push to remind to include the harms*
 - *Serious less frequent harms > frequent less serious harms*
 - *Frequency and seriousness.*
 - *There is scientific interest in a harm that is unique to a certain intervention or treatment that there is interest to know more about.*
 - *Public health impact and downstream cost (financial and costs ensuing g from other treatments, opportunity cost).*
6. Do you use GRADE or another formal process to prioritize which harms to review? If so, what has been your experience using such methods?
- *Implement use of decision making framework that considers bringing both expert opinion and the systematic assessment at clinical trial level.*
 - *Guidance panel process: rate outcomes on a scale of 1-9. One being less important, 9 being most important.*
 - *Don't use GRADE or grading scale.*
 - *If trials classified an adverse event as primary or secondary outcomes, somebody has essentially decided it is important enough for them to collect the data.*
7. Does the type of evidence that is available for particular harms influence what types of harms should be selected for review? Do you prioritize data available from certain types of study designs (e.g., RCTs, cohort studies) or from certain types of sources (e.g., published literature vs. grey literature sources e.g. FDA or other regulatory agency). If so, how?
- *Quality as well as type of supporting evidence, and effect size.*
 - *EMS and FDA*

- *Categorize something as unanticipated and then tell future researchers to look it over*
 - *Start with the harms, as the type of harm of interested in will determine which study design will answer that. Choose the right study designs that can be included in the review to answer that question.*
 - *Case reports have been utilized on occasion.*
 - *When designing a question, you may need different searches for each harm.*
8. Sometimes harms for an intervention may not be known before an analysis has been performed (e.g., MI with rofecoxib). Should reviewers routinely try to incorporate important unanticipated harms in their review?
- *When running into unanticipated harms, treat that a safety signal.*
 - *Identify if there is a knowledge gap before taking any regulatory actions.*
 - *First need to ask what the priorities are: then type of info in study designs that may yield it should be the driver.*
 - *If you spot it in the literature and it looks serious it should be incorporated somewhere; at lease let people know it has been reported.*
 - *Reviewers should have an open mind about these things; if something is suddenly found in 3-4 studies; there should be some way to incorporate into protocols or reviews*
 - *Recommend trying to identify and retrieve data from other sources whenever possible.*
 - *It should be a priority*
 - *It's fair to report that and share that caution but better to let readers know about these harms even if not thought of.*
9. How should input from technical and clinical experts and patients or patient advocacy groups be used in selecting and prioritizing harms for review?
- *Patient incorporation has been a large focus.*
 - *Gather survey data for guideline development group*
 - *Well done surveys add something significant from the persons affected by the recommendation.*
 - *For patients, it depends on the conditions and the efficacy groups*
 - *There could be a panel of people who understood to assess outcomes and you could call on them in times if you need more condition specific input.*
 - *Input would improve over time.*
 - *Engaging patients through surveys was challenging for various reasons.*
10. What other sources are useful for determining which harms should be included in a review?
- *EMS and FDA*
 - *Regional databases for intervention effectiveness and programmatic data*
 - *Pharmaceutical companies. How can we get them to share data and grey literature?*
 - *Should consider engaging industry in key informants*
 - *Specially emphasizing talking to patients.*
11. Are there situations where the selection and prioritization of harms may change during the course of a review? What types of situations would it be reasonable to select different or additional harms for review?
- *Don't always get the PICO right at the beginning and we don't have extensive stakeholder engagement processes that the EPC programs do*

12. Do you think systematic reviews adequately explain how harms have been selected for review? What types of information should be provided for users/readers of the review with regard to selection and prioritization of harms?

- *Seriousness of harms, frequency, quality of the evidence, consistency, and perspectives from both physicians and patients on how to effectively mitigate risks, or prevent risks*
- *Need to hold harms as the same rigorous effects as benefits.*
- *Planning proposals should drive people to think about harms from the get-go.*
- *Explicitly stating the prioritization process, stakeholder involvement if any, stating these things would be helpful to readers to understand what are you actually*
- *At the minimum report rationale for choosing these harms and then the resources used.*
- *Being transparent, on everything is the most useful. There's a call for guidance transparency but we also need that in the reviews we base them on.*

Appendix E. Key Informants

Stanley Ip, M.D.

Patient-Centered Outcomes Research Institute (PCORI)

Yoon Loke, MBBS, MRCP, M.D.

Cochrane Adverse Effects Methods Group

Hassan Murad, M.D.

Mayo Clinic Evidence-Based Practice Research Program

Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group

Susan Norris, M.D., MSc, MPH

World Health Organization (WHO)

Melissa Starkey, Ph.D.

American College of Physicians

Cunlin Wang, M.D., Ph.D.

US Food and Drug Administration (FDA)

Appendix F. Review of EPC Reports

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
<p>Imaging Tests for the Staging of Colorectal Cancer</p> <p>ECRI Institute—Penn Medicine, 2014</p>	<p>Adult patients (> 17 years)</p>	<p>What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management? What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?</p>	<p>Endoscopic rectal ultrasound (ERUS), CT, MRI, (PET)/CT</p>	<p>ERUS: pain; minor bleeding</p> <p>CT/PET: radiation exposure (CT-10mSv of radiation, PET/CT-18mSv); MRI- allergic reactions to intravenous contrast agents and nephrogenic systemic fibrosis</p>
<p>Imaging Tests for the Diagnosis and Staging of Pancreatic Adenocarcinoma</p> <p>ECRI Institute—Penn Medicine, 2014</p>	<p>Adults with symptoms of: suspected pancreatic adenocarcinoma; established diagnosis of pancreatic adenocarcinoma;</p> <p>Adults without symptoms with high risk of having pancreatic adenocarcinoma.</p> <p>Had to report data from groups of in which at least 85% of the patients were from one of the patient populations of interest. If a study reported multiple populations, it must have reported data separately</p>	<p>What are the rates of harms of imaging techniques (e.g., MDCT, MDCT angiography, EUS-FNA, PET/CT, MRI) when used to diagnose and/or stage pancreatic adenocarcinoma? How are patient factors related to the harms of different imaging techniques? What are patient perspectives on the tolerance of different imaging techniques and the balance of benefits and harms of different imaging techniques?</p>	<p>Multidetector computed tomography (MDCT), PET/CT, endoscopic ultrasound with fine needle aspiration (EUS-FNA), MRI</p>	<p>MDCT and PET/CT: cancer causing radiation; EUS-FNA: pancreatitis; pain; puncture; perforation; bleeding.</p> <p>MRI reactions to contrast media</p>

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
Therapies for Clinically Localized Prostate Cancer: Update of a 2008 Systematic Review ECRI Institute—Penn Medicine, 2014	Prostate cancer	Common and severe adverse events of biopsy and treatment	Radical prostatectomy (RP); radiation therapy	RP: Urinary incontinence; erectile dysfunction; bowel dysfunction. Radiation therapy: genitourinary toxicity; gastrointestinal toxicity Biopsy: bleeding; nosocomial infection
Combination Therapy Versus Intensification of Statin Monotherapy: An Update Johns Hopkins University, 2014	Adults with moderate or high cardiovascular disease risk	Compared with higher dose statins and to one another, do combination regimens differ in benefits and harms within subgroups of patients?	Dual agent therapy (statin + lipid modifying medication) and intensification of statin therapy	
Cardiac Troponins Used as Diagnostic and Prognostic Tests in Patients With Kidney Disease Johns Hopkins University, 2014	We included studies of adult patients with CKD including ESRD. All studies included human subjects exclusively. Included patients who also are clinically suspected of having ACS.	What are the harms associated with a false-positive diagnosis of ACS based on an elevated troponin level?		No studies reported harms on harms associated with a false positive

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
Diagnosis of Right Lower Quadrant Pain and Suspected Acute Appendicitis Brown Evidence-based Practice Center, 2015	Patients with acute RLQ abdominal pain (≤7 days duration) for whom appendicitis was considered in the differential diagnosis. Separate analyses were performed for children (age <18 years); adults (age ≥18 years); women of reproductive age, pregnant women, and the elderly	What are the harms of diagnostic tests per se, and what are the treatment-related harms of test-directed treatment for tests used to diagnose RLQ pain and suspected acute appendicitis?	Surgery: laparoscopic or open appendectomy Testing: ionizing radiation from CT scans	Contrast related: vomiting after oral contrast; mild skin rash. Low risk for Gastrografin-induced chemical pneumonitis, leakage colonic contrast material, nausea, vomiting, extravasation of intravenous contrast material. Surgery related harms: Low risk for complications with laparoscopy (<10%). Higher rate of intra-abdominal abscess with laparoscopy (3.9%) than with open appendectomy
Treatments for Fibromyalgia in Adult Subgroups Minnesota Evidence-based Practice Center, 2015	Adults over 18; Compared treatments for fibromyalgia in subgroups of adults who were followed 3 months or longer after treatment initiation.	What are the harms of treatments for fibromyalgia in each of these specific adult subpopulations?	Pharmacologic (raloxifen, transdermal 17β-estradiol, duloxetine, naltrexone); exercise: deep water running	Deep vein thrombosis; leg cramps; anxiety; flushing; nausea; headache; dry mouth; vivid dreams; insomnia; muscle pain (exercise related)
Decisional Dilemmas in Discontinuing Prolonged Disease- Modifying Treatment for Multiple Sclerosis Minnesota Evidence-based Practice Center, 2015	Multiple sclerosis	What is the evidence for long-term harms [in discontinuing disease modifying treatment (DMT)]?	Disease modifying treatments (short term/ long term)	Harms for injectable DMTs do not differ between short and long term; most discontinuation occurs in short term; insufficient evidence for whether rebound after natalizumab exists or risk of fetal exposure to DMT

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
<p>Treatment of Nonmetastatic Muscle-Invasive Bladder Cancer</p> <p>Pacific Northwest Evidence-based Practice Center, 2015</p>	<p>Adults with node-negative, non-metastatic muscle-invasive bladder cancer. This includes TNM staging of T2, T3 or T4a, N0, M0.</p>	<p>What are the comparative adverse effects of various treatments for non-muscle-invasive bladder cancer, including intravesical chemotherapeutic or immunotherapeutic agents and TURBT? What are the comparative adverse effects of various tests for diagnosis and post-treatment surveillance of bladder cancer, including urinary biomarkers, cytology, and cystoscopy?</p>	<p>Chemotherapy; radiation therapy; and radical cystectomy, with or without regional lymph node dissection</p>	<p>Mortality; recurrence of bladder cancer; progression or metastasis of bladder cancer; quality of life; functional status</p>
<p>Emerging Approaches to Diagnosis and Treatment of Non-Muscle-Invasive Bladder Cancer</p> <p>Pacific Northwest Evidence-based Practice Center, 2015</p>	<p>Non-Muscle-Invasive Bladder Cancer</p>	<p>What are the comparative AE's of treatments for NIMI Bladder Cancer? How do AE's of treatment vary by patient characteristics, such as age, sex, ethnicity, performance status, or medical comorbidities</p>	<p>Local resection with transurethral resection of the bladder tumor (TURBT), often with adjuvant intravesical therapy to destroy residual tumor cells using chemotherapeutic agents (e.g., mitomycin C [MMC], apaziquone, paclitaxel, gemcitabine, thiotepa, valrubicin, doxorubicin, epirubicin), bacillus Calmette-Guérin (BCG), or interferon immunotherapy</p>	<p>Bladder cancer recurrence; bladder cancer progression; all-cause mortality; bladder cancer mortality; and local and systemic adverse events</p>

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
<p>Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome</p> <p>Pacific Northwest Evidence-based Practice Center, 2015</p>	<p>Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome</p>	<p>What harms are associated with diagnosing ME/CFS? What are the harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?</p>	<p>Graded Exercise Treatment; pharmaceutical; counseling/behavioral therapy; complimentary alternative medicine</p>	<p>Diagnosing: psychological harms; labeling; risk from diagnostic tests; and misdiagnosis. Therapeutic: suppression of adrenal glucocorticoid responsiveness; increased appetite; weight gain; difficulty sleeping with hydrocortisone; flu-like syndrome; chills; vasodilatation; dyspnea; dry skin with rintatolimod; headaches with immunoglobulin G; discontinuation of treatment with fluoxetine; nephrotoxicity with acyclovir</p>
<p>Health Information Exchange</p> <p>Pacific Northwest Evidence-based Practice Center, 2015</p>	<p>Health Information Exchange (HIE) electronic sharing of clinical information across the boundaries of health care organizations</p>	<p>What harms have resulted from HIE? Do harms vary by type of HIE? Do harms vary by health care settings and systems? Do harms vary by the IT system characteristics? How does the usability of HIE impact effectiveness or harms for individuals and organizations? What specific usability factors impact the effectiveness or harms from HIE?</p>		<p>Harms of HIE insufficiently studied</p>

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
<p>Nonpharmacological Versus Pharmacological Treatments for Adult Patients With Major Depressive Disorder</p> <p>RTI International– University of North Carolina Evidence-based Practice Center, 2015</p>	<p>Major Depressive Disorder/ adult patients</p>	<p>In adult patients with MDD, what are the comparative risks of harms of these treatment options: (1) for those undergoing an initial treatment attempt or; (2) for those who did not achieve remission following an initial adequate trial with an SGA? Do the comparative risks of treatment harms vary by MDD severity? Do the benefits and risks of harms of these treatment options differ by subgroups of patients with MDD defined by common accompanying psychiatric symptoms (coexisting anxiety, insomnia, low energy, or somatization) or demographic characteristics (age, sex, race, or ethnicity)?</p>	<p>Pharmacotherapy; psychotherapy; complimentary alternative medicine; exercise</p>	<p>Low to insufficient risk overall of remission; treatment discontinuation; drug interaction; adverse events across all treatment interventions</p>
<p>Management and Outcomes of Binge-Eating Disorder</p> <p>RTI International– University of North Carolina Evidence-based Practice Center, 2015</p>	<p>Individuals of all races, ethnicities, and cultural groups in one of three subpopulations: (1) meeting DSM-IV or DSM-5 criteria for BED; (2) post bariatric surgery patients with LOC eating; or (3) children (6 years to adolescence) with LOC eating. Because LOC eating has no commonly accepted definition, studies included in the review may define LOC eating using different diagnostic criteria</p>	<p>What is the evidence for harms associated with treatments for binge-eating disorder? What is the evidence for harms associated with treatments for loss-of-control eating among bariatric surgery patients? What is the evidence for harms associated with treatments for loss-of-control eating among children?</p>	<p>Pharmaceutical; psychological; behavioral</p>	<p>Pharmaceutical: SNS arousal; GI upset; sleep disturbance; insomnia.</p> <p>Behavioral: limited evidence of harms. Psychological: none reported</p>

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
Behavioral Programs for Diabetes Mellitus University of Alberta, 2015	T1DM, included studies of patients (any age) diagnosed with T1DM and who had undergone basic diabetes education. For T2DM, we included studies of adults with T2DM who had undergone basic diabetes education	For patients with T1DM, what are the associated harms (i.e., activity-related injury) of behavioral programs implemented in a community health setting compared with usual care, standard care, or active comparators?	T1DM: behavioral programs T2DM: multicomponent programs	No studies reported harms on activity related injuries from behavioral interventions
Treatments for Ankyloglossia and Ankyloglossia With Concomitant Lip-Tie Vanderbilt Evidence-based Practice Center, 2015	Ankyglossia and ankyglossia with concomitant lip-tie/ infants and children (0-18 years); studies with patients with Van der Woude syndrome, Pierre Robin syndrome or sequence, Down syndrome, or craniofacial abnormalities were excluded as were studies of premature babies (<37 weeks)	Harms of treatments for ankyglossia or ankyglossia with concomitant lip-tie	Surgery	Excessive bleeding; airway obstruction; pain; transient poor feeding secondary to discomfort; dysphagia; complications related to dysphagia such as aspiration pneumonia; surgical site infection; nerve damage; salivary gland damage; ranulae; scarring; soft tissue damage; oral aversion; re-adherence; need for further surgery/revision
Management of Postpartum Hemorrhage Vanderbilt Evidence-based Practice Center, 2015	Women with Postpartum Hemorrhage (PPH) immediately post-birth to 12 weeks postpartum following pregnancy of > 24 weeks' gestation	What are the harms, including adverse events, associated with interventions for management of postpartum hemorrhage?	Pharmacologic; Procedural; Surgical	Thrombotic complications; infertility; PPH in subsequent pregnancy; spontaneous abortion in subsequent pregnancy; hematoma; ureter lesions; reoperation; infection; bladder lesion

Report Title EPC, Year	Condition/ Population	Harms Related Key Question(s)	Interventions	Harms Outcomes
Psychosocial and Pharmacologic Interventions for Disruptive Behavior in Children and Adolescents Vanderbilt Evidence-based Practice Center, 2015	Disruptive behavior disorders/ Children under 18	What are the harms associated with treating children under 18 years of age for disruptive behaviors with either psychosocial or pharmacologic interventions?	Psychosocial; pharmacologic; combined	