

Challenges in Conducting EPC Reviews of Behavior Change Interventions



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

Challenges in Conducting EPC Reviews of Behavior Change Interventions

Prepared for:

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

Contract No.: 290-2012-00004-C

Prepared by:

Scientific Resource Center
Portland, OR

Investigators:

Melissa McPheeters, Ph.D., M.P.H.
Mary Butler, Ph.D.
Margaret Maglione, M.P.P.
Meera Viswanathan, Ph.D.
Maya O'Neil, Ph.D.
Richard Epstein, Ph.D., M.P.H.
Johanna Anderson, M.P.H.
Jeanne-Marie Guise, M.D., M.P.H.
Aysegul Gozu, M.D., M.P.H.
Karli Kondo, Ph.D., M.A.
Christine Chang, M.D., M.P.H.

This report is based on research conducted by Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers' 2014 Methods Workgroup 1. 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).

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.

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

This research was funded through contracts from the Agency for Healthcare Research and Quality to the following Evidence-based Practice Centers: Vanderbilt University (290-2012-00009-I), University of Minnesota (290-2012-00016-I), RAND Corporation (290-2012-00006-I), RTI International and the University of North Carolina (290-201200008-I), and the Scientific Resource Center for the EPC Program (290-2012-00004-C).

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

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

Suggested citation: McPheeters M, Butler M, Maglione M, Viswanathan M, O'Neil M, Epstein R, Anderson A, Guise J-M, Gozu A, Kondo K, Chang C. Challenges in Conducting EPC Reviews of Behavior Change Interventions. Methods Research Report. (Prepared by the Scientific Resource Center under Contract No. 290-2012-00004-C). AHRQ Publication No. 15-EHC029-EF. Rockville, MD: Agency for Healthcare Research and Quality; August 2015. 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 they 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 and 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.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

David Meyers, M.D.
Acting Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

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

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

Challenges in Conducting EPC Reviews of Behavior Change Interventions

Structured Abstract

Objective. This report describes the challenges faced by systematic reviewers in the Evidence-based Practice Center (EPC) program when reviewing behavior change interventions, and considers whether current guidance is specific enough to address these challenges.

Methods. A workgroup of members from EPCs, the Scientific Resource Center (SRC), and AHRQ was convened to describe systematic review methods for behavior change interventions in the EPC program and similar programs, and identify gaps in guidance. Complementary methods including a literature scan and key informant interviews were undertaken to reach the objectives. A literature scan was conducted to identify current guidance and methods literature on the inclusion of behavior change literature in systematic reviews. Interviews were held with thought leaders in the field to identify current practices and opinions on using behavior change literature. Workgroup members summarized information from the literature and interviews.

Results. We identified specific challenges related to study selection (particularly in using single subject experimental designs), data extraction, risk of bias, strength of evidence, presentation and quantitative analysis. Challenges are particularly acute when fields are focused on identifying and describing heterogeneity in populations and treatment effects rather than identifying generalizable effects. Guidance is available, but the sources of guidance most used by the EPC program (EPC, Cochrane, and the National Institute for Health and Care Excellence) lack specificity in addressing these challenges.

Conclusions. Challenge of reviewing behavior change literature exists at each step of EPC systematic reviews. A larger discussion is needed about how EPC program methods can meet the evidentiary needs and questions of decisionmakers, consistent with the program's aim to provide accurate, independent, and scientifically rigorous information. Any methods approaches moving forward should address the specific needs of decision makers for using reviews of behavior change literature.

Contents

Background	1
Objectives	1
Rationale	1
Methods	4
Overview.....	4
Literature Scan.....	4
Review of AHRQ EHC Reviews.....	5
Key Informant (KI) Interviews.....	5
Results	6
Overview.....	6
Specific Methodologic Challenges Faced by Investigators Conducting EPC Reviews of Behavior Change Interventions and Available Guidance.....	7
Scope and Key Questions	7
Study Selection	7
Data Extraction	9
Risk of Bias Assessments	11
Synthesis	12
Strength of Evidence.....	13
Presentation of Results and Conclusions, Including Applicability.....	15
Summary and Implications	17
Future Directions	18
References	20
Figures	
Figure 1. Behavior Change for Individuals.....	3
Appendixes	
Appendix A. AHRQ EPC Reports Including Studies of Behavioral Interventions	
Appendix B. Approaches to Addressing Behavior Change Interventions in EPC Reviews	
Appendix C. Key Informants	
Appendix D. Key Informant Interview Guide	

Background

Objectives

This white paper is the result of a scoping exercise that describes the challenges faced by systematic reviewers in the Evidence-based Practice Center (EPC) Program when reviewing behavior change interventions, and considers whether current guidance is specific enough to address these challenges. This lays the groundwork for consideration of next steps in methods development.

Rationale

For the purposes of this white paper, behavior change interventions are defined as interventions intended to change behavior at the individual level for the purposes of improving a health outcome. End-users of systematic reviews of behavior change interventions generally seek an improved understanding of which interventions work, for whom, and under what conditions.¹ In other words, the most useful systematic reviews address both causality and context: they place equal emphasis on whether an intervention works for anyone and in whom and under what circumstances it has an effect. Even for interventions that rely on explicitly tested biological mechanisms and do not require individual behavior change other than adherence to medication plans, effectiveness can be altered by variations in the dose, duration, and timing of the intervention.¹ Systematic reviews for biomedical interventions generally control for rather than explain these sources of variation.² Behavior change interventions include much greater complexity and therefore have many more sources of variation than biomedical interventions. Thus, behavior change interventions raise particular challenges in terms of answering questions of context-specific effectiveness.

First, complexity is magnified significantly for behavior change interventions that rely on complex causal mechanisms requiring several intermediate changes before health outcomes improve. Second, actors at multiple levels (individuals, providers, systems, and policy) may all influence behavioral change: studies vary in the degree to which they account for the interplay among these actors and settings. Third, the longstanding debate on the relative weight and importance of these agents of change at different levels in influencing behavior change influences the types of designs and resultant information available to synthesize.^{3,4} Studies that focus on individual agency may use traditional experimental or quasi-experimental designs that are amenable to systematic reviews but produce relatively little information on the interplay between intervention and context. Studies that focus on context (often including qualitative components or qualitative design) may be limited in their ability to produce generalizable information. Fourth, studies do not always articulate or report on the components and characteristics of the intervention arm and any control arms, the theorized causal mechanism, and the fidelity of the intervention as implemented. Finally, the tools available to systematic reviewers (largely quantitative) to deal with heterogeneity are limited in their ability to explain context-specific effectiveness when primary studies do not articulate or report on mechanisms of change.

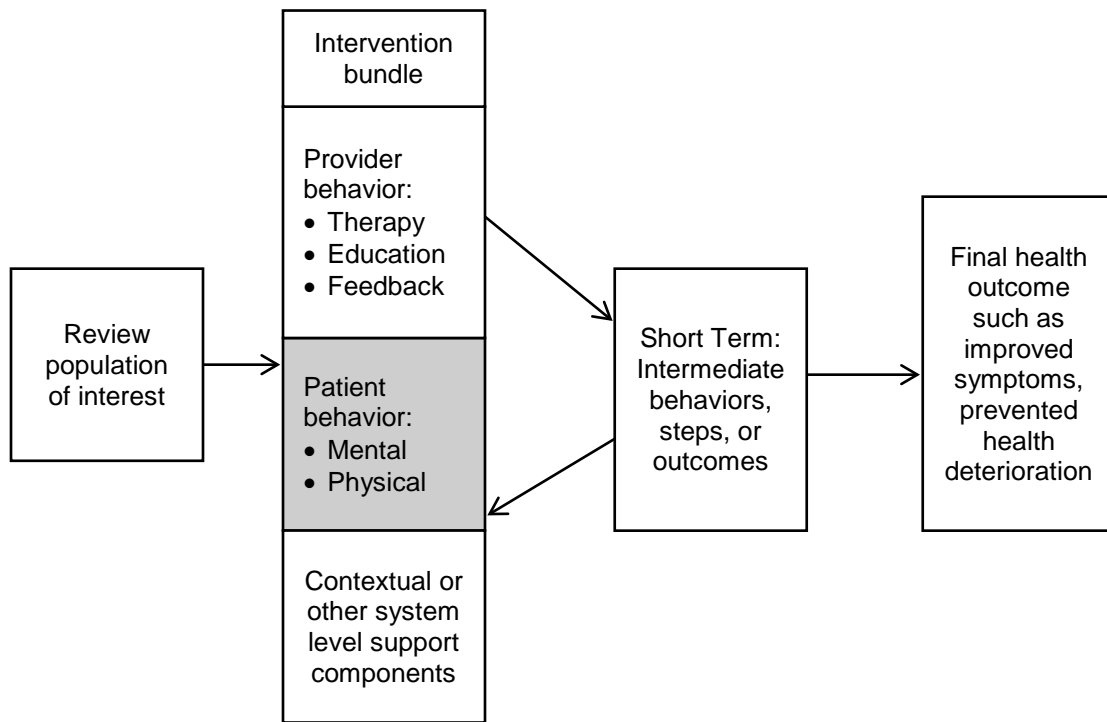
We further refine behavior change interventions for this paper to specifically include interventions intended to change patient behaviors. Thus, in addition to the complexity of interventions that often involve multiple components and may require change at the health services level, patient behavior change interventions can be seen as incorporating the patient as

part of the intervention. In order to achieve the final health outcome targeted, such as reduction in mental health symptoms or avoiding diabetic complications, patients themselves must change the behaviors that contribute to worsening health states, or add behaviors that may improve health states, and sustain those changes over time. Behavior change for an individual is actually part of the intervention that leads to health outcomes and can be measured as “process” or intermediate outcomes (Figure 1). The mechanisms of change for individual behavior change, then, would include the psychological processes that interact within an individual to produce sustained behavior change. Accordingly, while a recent EPC methods paper published on systematic reviews of complex multi-component interventions² and the on-going efforts to develop methods for complex interventions overlaps with and informs our discussions, our focus on patient behavior change interventions provides a unique contribution. However, a similar framework, and more familiar, could be constructed with an intervention box comprised of clinic-level behaviors (patient-centered medical home, for example), individual provider behaviors (behavior changes specific providers are required to adopt), and other contextual factors.

Also concurrent with this project, a different workgroup is developing guidance for the potential contributions of non-randomized studies (NRS) to comparative effectiveness reviews, and how NRS should be treated in such reviews. Behavior change interventions are very often assessed with NRS designs. At the clinic level, NRS designs often include pre/post designs with or without comparators or interrupted time series. For patient behavior change interventions, the analogous study design is a single-subject research design (SSRD), although SSRDs can be either randomized or not randomized. This paper will explicitly examine implications of this study design for patient behavior changes.

This workgroup was convened to consider whether there are methodologic challenges that are unique to, or more salient to, reviews of behavior change, and if so, whether existing EPC guidance documents adequately provide guidance for conducting systematic reviews of behavior change interventions. Each of the workgroup members has worked on reviews previously in which experts in behavioral research have suggested through peer or public review that the EPC methods are necessarily too limited to appropriately gauge the effectiveness of the primary literature in their field. We sought to determine whether this was true and whether any aspects of guidance warranted expansion or clarification.

Figure 1. Behavior change for individuals



Methods

Overview

A workgroup of members from EPCs, the Scientific Resource Center (SRC), and AHRQ was convened to describe systematic review methods for behavior change interventions in the EPC and similar programs, and identify gaps in guidance. Workgroup members participated in bimonthly teleconference calls to discuss project direction, assign tasks, and review draft documents. Complementary methods including a literature scan and key informant interviews were undertaken to reach the objectives. A literature scan was conducted to identify current guidance and methods literature on the inclusion of behavior change literature in systematic reviews. Interviews were held with thought leaders in the field to identify current practices and opinions on using behavior change literature. Workgroup members summarized information from the literature and interviews.

Literature Scan

To identify literature that addressed methods for including studies of behavior change interventions in systematic reviews, we searched multiple databases including MEDLINE via PubMed, PsycInfo, and ProQuest. We used a series of iterative strategies in each database including controlled vocabulary and key word terms such as research design, evidence-based medicine/methods, epidemiologic methods, and systematic review; and searches of journals that report methodologic studies (e.g., *Journal of Clinical Epidemiology*, *Behavioral Research Methods*) along with key words such as methodology or protocol following items related to particularly relevant studies identified in the searches. We searched gray literature to identify guidance from organizations completing systematic reviews of behavioral studies. Organizations include the Cochrane Collaboration, the Campbell Collaboration, US Preventive Services Task Force (USPSTF), Community Preventive Services Task Force, and the NICE. We also searched the EHC library of methods-related papers and guidance and the list of Effective Health Care (EHC) Program published reviews.

Investigators reviewed the retained materials for coverage of methodologic topics specific to the appraisal and synthesis of behavior change interventions. Data were extracted for each of the retained references to indicate methodologic topic coverage and record descriptions. We used a broad set of criteria to identify existing guidance and other relevant materials from the search results. We retained and categorized results as manuals, guidance, handbooks, or review methodology, with special attention to the materials authored or adopted by entities that assess behavior change interventions. We also retained supplemental publications that described the development or provided a detailed description of the guidance.

We conducted a preliminary scan of AHRQ EHC reviews to identify example reports that address studies of behavioral-based interventions (Appendix A). Two staff members reviewed the key questions, scope, and methods and results of each to identify relevant reviews. We recorded whether or not each review included behavior change interventions and if those assessments were described in the methods section.

Review of AHRQ EHC Reviews

The workgroup identified exemplar AHRQ EHC reviews that described various approaches to challenges of behavior change intervention reviews. Approaches to addressing behavior change interventions in reviews were summarized and are available in Appendix B.

Key Informant (KI) Interviews

We conducted open-ended interviews with six key informants for this project. Five were affiliated with universities in the United States or UK; one worked for the Centers for Disease Control and Prevention (CDC). One was chair of the methods group of the Campbell Collaboration at the time of the interview. All had several years of experience conducting systematic reviews of public health, behavior change, or educational interventions. Each KI completed an “EPC Conflict of Interest Disclosure Form” prior to being interviewed and no disclosed conflicts precluded participation of any of the invited organizations. A list of the key informants is provided in Appendix C. An interview guide was developed and used to prompt discussion during the interviews (Appendix D). The interview guide covered practices and challenges in conducting reviews of behavior change interventions including which study designs can be included, when to focus at the population versus the individual level, which questions should be answered by reviews of behavior change interventions, and whether or not behavioral interventions can be used to establish effects in systematic reviews. Investigators interviewed KIs about their experiences in producing reviews of behavior change interventions. All interviews were digitally recorded and transcribed and the transcripts were reviewed to identify and code key themes using NVivo software. The interviews focused on how and when these experts incorporated studies of behavior change interventions into systematic reviews. In addition, two KI interviews for another EPC Program Methods Workgroup on integrating bodies of evidence touched on relevant issues, so we included their feedback where appropriate.

Results

Overview

We sought to identify key challenges to conducting reviews of behavioral interventions under the EPC Program, and to assess whether the sources of guidance typically accessed by EPC investigators hold recommendations to address these challenges. We conducted a literature search as described above, and consulted with key informants. In the literature search, we identified 243 unique references. Of these, we excluded 142 from further analysis because they did not address behavior change intervention research or methodologic issues related to behavioral-based interventions. From the set of 88 relevant papers, we categorized 27 as handbook or guidance; the remaining documents were categorized as research (27) or review articles (34). We selected the subset of these that would most likely inform the practice of EPCs for analysis below in the text, but provide an overview of the totality of what is available in the Appendix.

Documents ranged from publication guides, such as the extension of the CONSORT guidelines, to methods used by professional societies for their reviews and guidelines,⁵ to examples of detailed methods sections used at organizations that review behavior change interventions.⁶

Relevant guidance exists in the Cochrane Handbook, the USPSTF Methods Guidance, and in the NICE Guidance, although none of it claims to speak directly to the issues of behavior change interventions as distinct from other interventions. Of note, the Cochrane Handbook, particularly recognizes the challenges in conducting reviews that are almost exclusively or exclusively composed of non-RCTs, noting that these are “much more difficult than carrying out a systematic review of randomized trials.”⁷ The Cochrane Handbook sections on Special Topics is of particular use to the reviewer of behavior change interventions, as it includes guidance on the use of non-randomized trials, patient-reported outcomes, qualitative research, and public health and health promotion.

The USPSTF provides perhaps the most direct discussion of methods in their document on how to assess behavioral counseling research. They identify several key challenges to applying USPSTF systematic review methods, which overlap somewhat with the challenges our team identified. These include:

- Study populations that may be inadequately characterized to assess risk or that are extremely heterogeneous
- Need to assess feasibility issues, including ease of enrollment
- Outcomes and assessment issues, including the challenge of identifying adverse effects of behavioral interventions, and the use of multiple outcome measures to assess a common construct

Although the guidance recognizes these challenges, no solutions are provided for systematic reviewers. Instead, the recommendations sections focuses primarily on making recommendations to behavioral researchers in terms of how to conduct and report their research to make it more amenable to reviews.

Specific Methodologic Challenges Faced by Investigators Conducting EPC Reviews of Behavior Change Interventions and Available Guidance

Below, we describe methodologic challenges as they pertain to study selection, data extraction, risk of bias assessment, strength of evidence and presentation, including applicability. These results reflect the experience of our workgroup, information from the literature and the thoughts of our key informants.

Scope and Key Questions

Challenges

As described above, the most relevant questions for stakeholders in reviews of behavior change interventions are often less focused on what interventions work at an overall population level, but rather which interventions work best under which circumstances for which individuals or target groups. Theoretical models often have been developed to make the case for an intervention approach and the intervention may be available, but may have suboptimal outcomes unless appropriately targeted, or appropriately modified for fit to an individual who is undertaking the behavior change. Thus, the key questions might need to be disproportionately focused on subgroup analyses in some reviews in the behavioral literature.

In the same way, many behavior change interventions are multi-component and complex, and it would be reasonable to conduct reviews either of the complex intervention's effectiveness overall, or to try and tease apart the degree to which individual components drive positive outcomes. This determination drives the choice of key questions and ultimately the scope of the review.

Available Guidance

The USPSTF procedure manual describes guidance for the inclusion of contextual questions within the scope of a review. These address important issues that require a valid assessment but not necessarily a systematic review. This information though is not reported as systematic results, but provide context and are described in the background and discussion and may be helpful in providing additional information on the circumstances in which certain interventions may be most effective.

Study Selection

Challenges

Study selection criteria typically are based on a predefined set of PICOTS, including for example pre-specifying a particular clinical population of interest. The full population serving as a target for a behavioral intervention may be defined more by characteristics and symptoms that cross clinical populations than by individual disease entities or syndromes, making searches broad and often unwieldy. Furthermore, the study design specification is often a challenge in reviews of behavior change interventions for the simple reason that there may be few RCTs or even comparative studies at all, with a frequent preponderance of single-subject design studies. The decision of whether to include and how to assess single-subject design research is complicated by a lack of guidance on how to do so. Because single-subject designs can take

many forms, and are commonly misunderstood in medical research, we described single-subject experimental designs (SSED) in some detail below. Finally, PICOTs for health system level behavior change interventions may suffer from “square peg in a round hole” syndrome, since the PICOT components alone may not capture the many pathways and processes associated with behavior change. Issues related to PICOT definitions may be better addressed on a case-by-case basis, but the EPC program can consider if broad brush stroke guidance, tips, or best practices would be useful.

Uncontrolled Studies

Observational studies without a control group are often included in EPC reviews of medical interventions in order to assess rare adverse events. RCTs are often too small to detect very rare events, and subjects who are more susceptible to adverse events due to age or pre-existing medical conditions are usually excluded from RCTs. However, including uncontrolled studies in systematic reviews to estimate efficacy is controversial. In behavioral research fields, RCTs may be difficult to implement for several reasons. Consumer pressure for access to a treatment may discourage the use of a control group. For example, it is most effective to intervene with children with ASD as soon as diagnosed;⁸ many parents would not tolerate their child being assigned to a wait list or no treatment.

Conference Abstracts

These summarize presentations and posters presented at scientific conferences. Such presentations do not undergo the rigorous peer review required by academic and professional research journals, and often do not include important information needed for EPC reviewers to assess the quality, applicability, and generalizability of the study. However, in fast-evolving fields, they may be the only results available on a specific new intervention or special population.

Inclusion of Single-Subject Experimental Design Studies

Single-subject experimental designs (SSED) have been used for the past 25 years to study behavior change interventions targeting children with various psychological, psychiatric, and physical conditions.⁹ Single-subject designs have recently been introduced for medical research; AHRQ published a user’s guide for the biomedical field in early 2014.¹⁰ In clinical medicine, so-called N-of-1 RCTs are multiple crossover trials, usually blinded and randomized, conducted in a single patient. More diverse single subject methods are used in the behavioral field, including reversal designs, multiple baseline designs, and designs with alternating treatments.¹¹ Importantly, SSED studies in the behavioral field usually involve more than one individual.¹² The intervention is applied to several subjects (usually fewer than 10) utilizing the same protocol and manual to maintain fidelity. In that way, researchers may isolate patient characteristics associated with improvement on an outcome.

Although not appropriate to determine overall effectiveness at the population level of an intervention, SSED have numerous important strengths. They may be conducted when an RCT is unfeasible because it is unethical to withhold treatment from a group. Also, they may be less expensive to conduct than RCTs and thus used to conduct important research when resources are lacking. Others use SSED based on the philosophical position that the best control for an individual patient is that same patient. In an era where patient-centered research is paramount, their focus on the effect of patient characteristics and specific intervention components may add

depth and breadth to AHRQ systematic reviews, which often include Key Questions on special populations, timing, dosage, and program characteristics. Still, if SSED studies are to be considered for future inclusion in AHRQ systematic reviews, several important issues must be addressed. One KI noted a current project to develop Bayesian methods to estimate whether N-of-1 studies approximate results from RCTs.

Two KIs had included single subject research in their SRs, with a particular focus on capturing details of contexts under which specific interventions could be effective; all KIs had included other non-RCT designs. One KI's SR on alcohol policy even included an ethnographic study. One informant used separate teams in parallel to review the RCTs and the non-RCTs and abstract data simultaneously. This is done to avoid any bias that might come from knowing the results of one set of studies earlier.

Available Guidance

SSED typically have been excluded from EPC systematic reviews due to issues of statistical power and generalizability to a wider population. In fact, a 2010 AHRQ-funded project testing an instrument to classify study designs did not include single subject studies.¹³ The section on non-randomized studies in the Cochrane Handbook includes guiding questions to help investigators classify studies into different designs, but is limited to the “classic” designs in biomedical literature, excluding, for example, stepped-wedge, time series or single-subject experimental design. There is no information available on how to include single-subject experimental design studies, including assessing risk of bias and incorporating into an overall synthesis. Similarly, the EPC methods guide on risk of bias assessment includes a taxonomy of study types, but does not include single-subject experimental design.

One KI reported that the next version of the Cochrane handbook will include a chapter on qualitative research, although it was unclear if, when, and under what circumstances this would be done.

The EPC Program described a “best evidence” approach for study selection for a review.¹⁴ This entails weighing the trade-offs in risk of bias, applicability, replication, conclusiveness, and overall strength of evidence. While potentially applicable to behavior change intervention reviews, offers a general approach without more specific considerations for study types that may be considered higher risk of bias.

Data Extraction

Challenges

Reviews of behavior change interventions often face a greater data extraction burden than reviews of drug interventions. At a very practical level this occurs, for example, when a drug can be described by a limited set of variables (e.g. name/chemical composition, dosage, frequency of administration) while a behavior change intervention may include many more components. This is a particular challenge when, as previously described, individual interventions are further tailored to the individual and therefore may have multiple forms or iterations. Unlike drug studies (but often similar to surgical or medical device interventions), the qualifications and skill level of the intervention provider(s) may also be important and should be captured in the data extraction. A study's attention to measuring fidelity to the intervention protocol should also be extracted including the specifics of how fidelity was assessed. Because of the additional information required to assess even simple single component behavior change interventions, full

journal articles reporting study design and results are generally required even to assess inclusion, eliminating the benefit of initially reviewing abstracts.

Furthermore, the number of outcomes of interest is often substantially higher in behavior change studies, including individual components of behavior change measures in addition to overall assessments. Target outcomes are frequently individually established. Reviewers report little success in limiting the number of outcomes of interest with the help of content experts. This is very likely a reflection of a point raised by KIs that there is frequently little consensus among experts as to which scales or measures are most appropriately used in any given behavior change field to assess important outcomes. Whereas in many medical interventions outcomes such as mortality or specific morbidities are objectively determined and dichotomous, outcomes of interest in behavioral research are often assessed by patient self-report questionnaires or behavioral observations, with continuous outcomes. The degree to which these outcomes are validated varies substantially.

In childhood research, for example, there are many validated instruments used to measure the same construct such as social skills or adaptive behavior. KI who raised the issue agreed that translating into an effect size for the purpose of meta-analysis was considered acceptable in the educational and psychology fields, provided that the results were translated into meaningful information for the practitioner.

Thus the data extraction process, particularly around description of the intervention and the outcomes, is frequently substantially greater in behavior change reviews than in biomedical reviews. Potentially greater time and effort may be needed to identify and specify interventions and their components can have significant impacts on timeliness. However, scaling down the scope of a review in response may affect the relevance and usefulness of a systematic review.

Available Guidance

The EPC program is currently working to establish minimum reporting elements to describe complex interventions, and these may be used to guide data extraction processes. The Cochrane Collaboration has also been attending to this issue. However, since behavior change interventions can occur at any one of several levels of a health system, without actual published material it is unclear whether such guidance will be sufficient to address SRs evaluating patient behavior change interventions.

The Cochrane Handbook and EPC Program methods guidance¹⁵ also addresses patient-reported outcomes, and recommends specific considerations when assessing outcome measures for a review. Both note that systematic reviewers need to be able to assess the reliability and validity of outcome assessment tools to assess the risk of biased results. This implies a basic understanding of questionnaire development and validity literature can provide the background needed to assess whether or not an outcome assessment tool is reliable and has been validated. However the guidance does not specify minimum expectations for a valid and validated measure for inclusion in reports. Guidance does not also specify the extent to which reviewers should confirm the validity of outcomes measures and does not require expertise in measure development and validity.

Risk of Bias Assessments

Challenges

Investigators on the team and KIs report an absence of established tools appropriate for assessing risk of bias in behavior change interventions. Underlying this lack of specific tools is disagreement in terms of what constitutes high quality behavior change research overall, including lack of consensus on which study designs are most appropriate (e.g. RCTs versus observational). As a result, there is some uncertainty as to whether the study design and rigor hierarchy accepted for biomedical research should be adopted across behavior research as well or whether a different model is best adopted. This includes disagreement about whether blinding is an essential element for risk of bias assessments in the behavior change literature and whether a behavior change study can have low risk of bias in the absence of blinding. This idea mixes concepts of study rigor with feasibility, reflecting the perceived difficulty of implementing good blinding procedures in behavior change studies.

Lack of Blinding

Bias can arise from lack of participant and personnel blinding to assignment and from lack of blinding of outcome assessors. For behavior change interventions, it may be difficult, and sometimes impossible, to blind participants and personnel because they often involve extensive patient and provider interaction and awareness of treatment components. Nonetheless, some outcomes such as objective cognitive assessments can be evaluated by blinded outcome assessors. For other outcomes such as self-reported symptoms, the risk of bias is compounded by the lack of patient or personnel blinding of assignment. The inherent difficulties of achieving patient or personnel blinding in behavioral interventions mean that many behavioral interventions may be rated as having a higher overall risk of bias than interventions that can more easily blind patients or personnel.

If blinding, regardless of feasibility or appropriateness, is an important measure of rigor, then by definition, many behavior change interventions can never be assessed with the highest rigor. Some SR approaches, such as that espoused by the US Preventive Services Task Force, exclude studies with a high risk of bias. In such cases, behavioral interventions may be disproportionately excluded from the body of evidence, resulting in instances where the review is unable to make a conclusion because of the lack of high-quality evidence. One must ask the question of “why bother” with a risk of bias assessment in which the assessment is dictated by the field, not by the study conduct. Certainly, some behavior change studies have used unique and creative methods of maintaining patient and provider blinding,¹⁶ these methods are often difficult to implement and are not commonly used. Thus, researchers are currently struggling with the question of whether key sources of bias in behavior change interventions center around blinding as often is the case in medical interventions, or whether other factors should carry more weight in assessing risk of bias. We did not find consensus on this issue.

Fidelity

Although the risk of performance bias, due to lack of fidelity, is not unique to behavioral interventions, behavior change interventions may include more intervention elements or intervention elements with a high degree of potential customization. In such interventions, the risk of performance bias may be elevated when fidelity to core and planned interventions is not monitored during the course of the intervention.

Rating Study Quality of SSED

Factors that increase internal validity of SSED include dual assessment of observed outcomes, inter-observer reliability, and collection of data at multiple time points in the absence or presence of the behavior change intervention. However, no quality rating instrument has been widely accepted. In 2008, Logan and colleagues published a 14-item instrument developed to rate the quality of an SSED study as strong, moderate, or weak.¹⁷ This instrument assesses aspects such as sufficient description for replicability, blinding of outcome assessment, and measure variability during the baseline period. The authors suggest that a high quality SSED will include at least three patients and five data points in each phase. Other important issues such as random allocation, loss to follow-up, and data missing at different time points are not addressed by the instrument. Logan et al. also proposed a five-level hierarchy of evidence for SSEDs, with N-of-1 RCTs at the top. In 2013, Tate and colleagues published the RoBiNT Scale, another instrument for rating quality.¹⁸ It includes 15 items in all, split in two subscales for internal and external validity, respectively. The first scale includes items on randomization, sampling, and blinding and the second scale focuses on aspects of thorough reporting and data analysis. Both the Logan et al. instrument and RoBiNT have displayed good construct validity and inter-rater reliability. Notably, both instruments go beyond traditional risk of bias (internal validity) concerns of instruments such as the Cochrane RoB tool to evaluate replicability¹⁷ external validity.¹⁹ SRs that use tool such as the Cochrane RoB instrument for trials and RoBiNT for SSED for the same evidence base will need to address the lack of concordance in underlying risk of bias criteria.

Available Guidance

None of the typically used sources of guidance provide risk of bias tools for use in the full range of study designs seen in the behavioral literature, although tools designed for cohort studies can often be adapted for this use. EPCs have typically developed their own approaches and use of the newest guidance on risk of bias does provide a bank of questions²⁰ that can be used for this purpose. Nonetheless, a determination of the applicability of approaches for assessing other study designs such as the SSED is currently unavailable.

Synthesis

Challenges

Although a detailed review of all the issues related to synthesis, including quantitative analysis, in reviews of behavior change interventions is beyond the scope of the current manuscript, one increasingly important issue involves the use of patient-reported outcomes. Behavior change interventions often involved the use of patient-reported outcomes. Patient-reported outcomes for behavior change interventions generally involve continuous measures and, for many outcomes, many different continuous measures exist purporting to assess the same (or at least similar) underlying constructs. Existing guidance recommends against combining measures of different underlying constructs, but does not make specific recommendations about how to make these determinations. Additionally, existing guidance recommends using the standardized mean difference (SMD) to combine different continuous measures of the same construct in meta-analysis, but also states that the use of SMDs may be problematic for this purpose in certain situations.^{7,21}

Complex or Multi-Component Interventions

In order to determine which intervention components are driving effectiveness and identify standardized effect measures, statistical techniques such as SEM or other multivariate and modeling analysis methods are often employed. Additionally, these types of analyses of multiple treatment components require statistical adjustment for multiple comparisons in order to ensure that statistical assumptions are met. While some of these analytic approaches to multi-component interventions are described in recent AHRQ publications, existing guidance does not specifically address analytical approaches applied to behavior change interventions,² leaving reviewers with little guidance on what data points to extract.

Available Guidance

The Cochrane Handbook examines the issue of multiple patient-reported outcome measures used for a particular construct that must be somehow synthesized. The guide states, “Investigators must ... fall back on intuitive decisions about the extent to which different instruments are measuring the same underlying construct. For example, the authors of meta-analysis of psychosocial interventions in the treatment of pre-menstrual syndrome faced a profusion of outcome measures, with 25 patient-reported outcomes reported in their nine eligible studies. They dealt with this problem by having two investigators independently examine each instrument – including all domains – and group them into six discrete conceptual categories; discrepancies were resolved by discussion to achieve consensus. The pooled analysis of each category included between two and six studies.”

A number of organizations and texts provide support for and guidance in quantitative synthesis of behavioral interventions, particularly traditional meta-analysis. The Campbell Collaboration, for example, emphasizes meta-analysis with a focus on calculating standardized effect sizes, given the heterogeneity of outcome measures commonly assessed in behavioral studies. Decisions about when and under what circumstances to combine studies quantitatively, particularly in the presence of heterogeneous outcome measures or populations may warrant some discussion at the program. The question may not just be how much heterogeneity is acceptable or not acceptable, but also how to identify the range of outcome responses across the populations so that effective responses are not washed out of the information.

Cochrane Chapter 9.2.3 on standardized mean differences provides guidance on the calculation of standardized mean differences, but concludes that methods of interpretation of SMDs remain unsatisfactory, although as we note, these approaches are often the best available for behavior change intervention reviews. They do specify that methods for synthesizing qualitative methods are evolving but not recommend at this time (e.g., meta-ethnography).

The EPC Program has embarked on methods work in integrating randomized and non-randomized bodies of evidence in a review. While it may be helpful to the program, it may not apply as directly to behavior change interventions.

Strength of Evidence

Challenges

Given the types and limitations of behavior change intervention research – including study designs for which the EPC program does not have a risk of bias tool consistently used across the program – incorporating behavior change studies into strength of evidence assessments is particularly challenging. Other challenges include the frequently substantial variation in the

interventions across the studies, and the wide range and number of individual outcome measures used, which complicate decisions about when to lump or split studies for SOE assessments.

Other organizations have developed SOE-type systems specifically describing recommendations for incorporating SSED, but none of these organizations is one that EPCs would be likely to refer to for systematic review methodology (see below). Therefore, while these methods exist, they have not been vetted for use in EPC reports and would be unlikely to be adopted (and indeed have not been in relevant reports).

For example, Horner et al.²² suggested the following criteria be met for a behavior change intervention investigated by SSED to be considered “established,” which is their highest classification:

“[1] a minimum of five SSED studies on the treatment have been published in peer-reviewed journals that meet minimally acceptable methodological criteria and document experimental control; (2) the studies are conducted by at least three different investigators across three different locations; and (3) the studies include a total of at least 20 participants.”

The National Autism Center used very similar criteria in its recent National Standards Report (NSR).⁶ An intervention may be rated as “established” if there were at least “4 single subject experimental design (SSED) studies with a minimum of 12 total participants for which there are no conflicting results” or “6 SSEDs with a minimum of 18 total participants with no more than one study reporting conflicting results” as long as these studies scored three or higher on a five point quality measure developed by the NAC. NAC classified 11 behavior change interventions for autism as established. These conclusions are in marked contrast with the recent AHRQ funded systematic review on the same topic, which found insufficient evidence of efficacy for all behavior change interventions except one early intervention model, classified as low SOE. As this example demonstrates, varying standards can lead to conflicting results in different reviews of the same topic, leading to a potential lack of confidence in SR methods and results.

The American Psychological Association²³ and the CDC provide guidance on incorporating the results of single subject research into SRs that included other study designs; however, there is no consensus in this area or validated method of rating Strength of Evidence (SOE) in such cases.

Available Guidance

Overall, guidance for questions that are associative rather than causal, and aimed at understanding context or context-specific effectiveness is not well developed. Strength of evidence techniques and guidance for questions answered by non-causal quantitative methods, or qualitative methods to gain insight into contextual issues are not articulated. Further, while not yet published, strength of evidence methodology for qualitative research, CerQUAL, was presented at a recent GRADE DECIDE conference in June 2014. The discussion for Question 2 of this paper also mentioned approaches and techniques for evaluating a body of evidence based on SSED.

The EPC Program recently published updated guidance for strength of evidence assessment, but does not describe particular considerations for different types of bodies of evidence, and how to translate strength of evidence grades for population-level research to individuals, local settings, systems, or healthcare providers.

None of the guidance documents examined included guidance about how best to evaluate or to include SSED in a strength of evidence assessment.

Presentation of Results and Conclusions, Including Applicability

Challenges

Because behavior change interventions can involve complex patients, complex interventions, or both, presenting results and conclusions can be challenging. In quantitative synthesis, effect measures are commonly standardized and lack immediate clinical interpretability. In many cases, quantitative synthesis is not possible, and with significant variation in studies and outcomes, tables and descriptions become long and may seem to lack synthesis. Since results may be organized in several reasonable ways, guiding the reader by providing general text outlining the organization becomes more important.

Applicability and Generalization

Studies used in behavior change studies often rely on convenience samples and a small number of patients.¹² Thus, the results are not generalizable to the entire population living with a particular condition and it may be reasonable to suggest that they may not be appropriate for obtaining overall estimates of effect. Applicability is difficult to assess. Indeed, the challenge is in defining what is a “representative population” for a particular behavior change intervention. SSEDs are a relatively extreme illustration of this problem. Some behavioral researchers suggest that SSED has so-called “logical generalizability” meaning that once SSED studies have been replicated many times, the results are applicable to patients with similar characteristics to those enrolled. This would, in turn, help describe the population the sample represents. This concept has not been validated empirically¹² but in theory, the SSED literature could be compiled to address subpopulations of special interest in systematic reviews. Of course, subject selection and characteristics would need to be described with replicable precision in the SSED articles.

Generalizability for SSEDs may also be conceived of as across settings for one individual rather than across individuals and if this approach is used, it should be clearly specified in systematic reviews. One KI suggested that SSEDs are best used when the goal is to understand how a behavioral intervention works for one individual. For example, an SSED may vary the settings in which the intervention is used – the clinic, at home, at school, at a grocery store – for evidence of the learning process underway and that the behavior change intervention is generalizable to multiple settings.

SSEDs are an extreme form of the broader issue of whether generalizability as generally conceived is a logical aim for conditions or syndromes that are highly heterogeneous. Such heterogeneity is often due to the multiple pathways through which a condition or syndrome develops. Or the heterogeneity may be due to the complexity of the existing patient characteristics present at the time the intervention is fielded that create the patient conditions within which intervention interacts to effect change. The presentation of results that don't conform to the standard of identifying a main effect size that combines the effects in individual studies is challenging. Reviews of behavior changes interventions are likely to have greater emphasis on subsets than a main effect, multiplying the need for the number of individual results presented in the text and complicating the ability to present a straightforward message. Applicability by patient patterns is not as easily incorporated by guideline developers into broad

guidelines, and may be more challenging for clinicians to use and apply. More particular findings, and the guidelines that would result, are more challenging to effectively communicate.

Available Guidance

The EPC methods guide indicates that applicability may depend on end-user needs, and the type of review question. Thus it recommends consultation with stakeholders to identify the most important factors that may affect applicability. The approach to assessing applicability in the EPC methods guide²⁴ includes using elements of the PICOTS to assess and judge applicability of individual studies and the body of evidence. These elements can inform judgment about differences in treatment outcomes, related to effect modification, variance in absolute benefits and harms, and generalizability to everyday practice. Factors affecting applicability and differences between the review evidence and ideally applicable evidence should be described to assist decisionmakers.

The Cochrane Methods Guide⁷ also notes the importance of identifying effect modifiers, differences in baseline risk, and identifying important variation across populations. Guidance recommends identifying groups with varying risk in the summary of findings table, and when making an assessment of applicability consider whether why evidence should not be applied to a particular patient. Important areas of variation that may limit applicability include biologic variation, context and culture, adherence, and values and preferences.

This guidance provides a general and flexible approach. Investigators can include other elements relevant to complexity seen with behavioral interventions. Forthcoming guidance related to complex multicomponent interventions may inform these efforts.

Summary and Implications

EPCs are conducting a growing number of reviews focused on behavior change literature. Behavior change interventions are used across a variety of health-related conditions, often in heterogeneous populations. Multiple outcomes may be studied, with outcomes individualized to the patient, and important outcomes may compete (for example, trading pain reduction for function). Interventionists often follow an approach that is theoretically and philosophically based, but modify the intervention to accommodate the needs and the interests of the individual, thus the interventions may be adapted to individuals within the studies. In addition, behavior change interventions are also intended to effect changes in how health care is provided to improve patient outcomes. These interventions are sometimes targeted to providers at the individual level, but may also involve system-level changes that aim to engineer systems that nudge or support individual behavioral changes. The nature of behavior change interventions make aggregation, synthesis, and communicating important findings of systematic reviews challenging.

Issues that create challenges in synthesizing behavior change literature arise in all main components of reviews, including setting the scope and key questions, study selection, data extraction, risk of bias, strength of evidence and presentation of results and conclusions. These issues are related in part to an overlapping but different set of study designs for which established guidance and methods are at times a poor fit (e.g. NRS), and for some even non-existent (e.g. SSED). There may be few RCTs or prospective studies available to review, and studies that are conducted may be small. Common study designs, analytic approaches, and the language used to describe behavior change intervention studies may be specific to the particular field and not clearly described in existing EPC methods.

In addition, many behavior change interventions are targeted to *symptoms*, rather than clinical conditions, and may be used and studied across a range of diagnosable conditions. Likewise, a particular symptom may be targeted by multiple behavior change interventions. The decision to focus on a symptom or an intervention and the range of conditions involved increases the complexity of the context and has implications for how to involve an appropriate range of stakeholders in the systematic review process. The prevalence of the use of multiple outcomes, which must be standardized for interpretation, adds to the difficulties. More fundamentally, though, the challenges arise from synthesizing bodies of literature focused not on obtaining average effects across generalizable populations, but on identifying and describing heterogeneity in treatment effects and populations.

Ultimately, the issue of whether useful information can be identified rests on strength of evidence. Under the traditional strength of evidence paradigms based on Bradford Hill guidelines, causal strength of evidence is determined by two factors – one is the confidence in the estimate of effect and the second is on the likelihood that additional research will change the findings. Low strength of evidence reflects limited confidence that the estimate of effect lies close to the true effect and that additional evidence is needed before concluding either that the findings are stable. In most cases, these two factors are in alignment. However, in the behavior health field, the study designs may limit the ability of the reviewer to conclude causality using only strong Bradford Hill guidelines, and future research is unlikely to provide any greater confidence, making it a challenge for decision-makers who must make decisions under such great uncertainty.

In addition, the question posed by a systematic review may be at odds with the focus of the studies themselves. For example, while policy makers may want to know about the overall

comparative effectiveness of two interventions, the individual studies available to answer that question are instead often designed to focus on explicating the context in which the intervention is most effective. Forcing these studies into a synthesis approach that seeks an overarching effect in generalizable populations may miss important contextual questions that are considered important in the field. A focus on individualized outcomes is contrary to the usual conclusions of RCTs in which inferences are made at a generalizable, group or average level. Some of the literature our team identified published by other investigators and review groups suggested that in evidence-based practice the usual study design hierarchy may be inappropriate for studying the effectiveness of some behavior change interventions and is, in fact, “upside down.”¹² In this view, the assumption that a sample does or should reflect a broad population with a defining clinical condition may be inappropriate and non-RCTs are preferred. Any differences observed in treatment effect are as likely to be about heterogeneity in the population as about treatment effects overall, especially in smaller studies, which are common in intensive, behavior change interventions. By the same token, a lack of observed effect may be due to heterogeneity masking a true effect in specific subgroups when an average effect is calculated. This philosophy is at odds with a system of risk of bias assessment and strength of evidence that prioritizes RCTs.^{15,25}

Given these limitations and uncertainty, the question remains as to whether systematic reviews can synthesize evidence in a way that is more helpful to decision-makers. Such an approach might place greater emphasis on research that provides contextual information, and may require use of a larger variety of study designs, and a different approach for assessing SOE.²⁶

While the current EPC guidance is intended to be applicable to all types of interventions, it does not specifically address differences in reviewing behavior change literature and does not provide advice on a number of issues, including an agreed upon approach to determine when populations and outcomes are similar enough to combine quantitatively, what additional measures of quality/risk of bias may be particularly important (e.g. fidelity), whether SOE grading can accommodate the need to distinguish between lower levels of evidence to help decision-makers. In the absence of guidance, methodological work in these specific areas may provide investigators more specific and relevant considerations for a more consistent approach for the program.

Future Directions

The Workgroup for Intervention Development and Evaluation Research (WIDER) issued recommendations to improve the reporting of behavior change interventions, specifically calling on studies to provide a detailed description of interventions in published papers, clarify the assumed change process and design principles, provide access to intervention manuals/protocols, and provide a detailed description of active control conditions.^{27,28} These recommendations, if widely adopted, will yield substantially greater amounts of relevant information and will help systematic reviews answer the question of context-specific effectiveness, perhaps with existing quantitative tools. The uptake of these recommendations is yet untested. In the meantime, end-users of systematic reviews seek actionable information now on behavior change interventions.

Our workgroup identified three main opportunities for additional methods work. First, investigators can explore opportunities to shape reviews to incorporate and focus on context and **“which interventions work best under which circumstances for which individuals or target groups.”** Second, investigators can pilot methods and approaches in EPC program systematic reviews. Finally, the program is preparing to launch methods work in complex interventions,

which is potentially applicable to behavior change interventions. This is an opportunity to leverage this work, and ensure applicability to a range of complex interventions, including behavior change interventions. As patient-centered outcomes research becomes more dominant in the field, the idea of synthesizing literature that focuses on describing heterogeneity more than reaching an overall estimate of effect is appealing but brings with it the challenges noted in this white paper. A decision to move in that direction requires examination of the methodologic issues raised here.

Key Observations

- Challenges of reviewing behavior change literature exist at each step of EPC systematic reviews.
- Challenges are particularly acute when fields are focused on identifying and describing heterogeneity in populations and treatment effects rather than identifying generalizable effects.
- A larger discussion is needed about how EPC program methods can meet the evidentiary needs and questions of decisionmakers, consistent with the Program's aim to provide accurate, independent, and scientifically rigorous information.
- Any methods approaches moving forward should address the specific needs of decision makers for using reviews of behavior change literature.

References

1. Glasziou P, Chalmers I, Altman DG, et al. Taking healthcare interventions from trial to practice. *BMJ*. 2010;341.
2. Guise J-M, Chang C, Viswanathan M, et al. Systematic Reviews of Complex Multicomponent Health Care Interventions. AHRQ Publication No. 14-EHC003-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2014. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
3. Connelly J. The Behavior Change Consortium studies: missed opportunities—individual focus with an inadequate engagement with personhood and socio-economic realities. *Health Education Research*. 2002;17(6):691-5.
4. Glasgow RE, Dzawaltowski DA, Estabrooks PA, et al. Response to Connelly. Response from the Behavior Change Consortium Representativeness and Translation Work Group: the issue is one of impact, not of world view or preferred approach. *Health Education Research*. 2002;17(6):696-9.
5. O'Donnell M, Darrah J, Adams R, et al. AACPD methodology to develop systematic reviews of treatment interventions. *American Academy for Cerebral Palsy and Developmental Medicine*; 2004. p. 1-30.
6. National Standards Report National Autism Center. Randolph, MA: National Autism Center; 2009. <http://mn.gov/mnddc/asd-employment/pdf/09-NSR-NAC.pdf>
7. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions* Version 5.1. 0 [updated March 2011] The Cochrane Collaboration. Cochrane Collaboration; March 2011. <http://handbook.cochrane.org/>
8. Taylor JL, Dove D, Veenstra-VanderWeele J, et al. Interventions for adolescents and young adults with autism spectrum disorders. AHRQ Publication No. 12-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; August 2012. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
9. Maggin DM, Odom SL. Evaluating single-case research data for systematic review: A commentary for the special issue. *J Sch Psychology*. 2014;52(2):237-41.
10. Kravitz RL, Duan N, The DEcIDE Methods Center N-of-1 Guidance Panel. Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
11. Single-case experimental designs. Workshop on Innovative Study Designs and Methods for Developing, Testing and Implementing Behavioral Interventions to Improve Health; 2014; Bethesda, MD.
12. Schlosser RW. The role of single-subject experimental designs in evidence-based practice times. *Focus*. 2009;22:1-8.
13. Hartling L, Bond K, Harvey K, et al. Developing and testing a tool for the classification of study designs in systematic reviews of interventions and exposures. AHRQ Publication No. 11-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2010. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
14. Treadwell JR, Singh S, Talati R, et al. A Framework for "Best Evidence" Approaches in Systematic Reviews. AHRQ Publication No. 11-EHC046-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2011. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
15. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. AHRQ publication No. 12-EHC047-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>

16. Hatcher S, Sharon C, Parag V, et al. Problem-solving therapy for people who present to hospital with self-harm: Zelen randomised controlled trial. *The British Journal of Psychiatry*. 2011;199(4):310-6.
17. Logan LR, Hickman RR, Harris SR, et al. Single-subject research design: recommendations for levels of evidence and quality rating. *Developmental Medicine & Child Neurology*. 2008;50(2):99-103.
18. Tate R, Perdices M, Rosenkoetter U, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological rehabilitation*. 2013(ahead-of-print):1-20.
19. Tate RL, Perdices M, Rosenkoetter U, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation*. 2013;23(5):619-38.
20. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epi*. 2012;65(2):163-78.
21. Methods guide for effectiveness and comparative effectiveness reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2014. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
22. Horner RH, Carr EG, Halle J, et al. The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*. 2005;71(2):165-79.
23. Thygeson M, Peikes D, Zutshi A. Mixed Methods: Fuzzy Set Qualitative Comparative Analysis and Configurational Comparative Methods: Powerful Methods to Study and Refine Patient-Centered Medical Home Models. AHRQ Pub No: 13-0026-EF Agency for Healthcare Research and Quality; February 2013. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
24. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of clinical epidemiology*. 2011;64(11):1198-207.
25. Berkman ND, Lohr KN, Ansari M, et al. Grading the strength of a body of evidence when assessing health care interventions for the effective health care program of the Agency for Healthcare Research and Quality: an update. AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2013. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
26. Howick J, Glasziou P, Aronson JK. The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? *Journal of the Royal Society of Medicine*. 2009;102(5):186-94.
27. Albrecht L, Archibald M, Arseneau D, et al. Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implement Sci*. 2013;8:52.
28. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci*. 2009;4(40):1-6.

Appendix A. AHRQ EHC Program Reports Including Studies of Behavior Change Interventions

Report	Examples of Behavioral Interventions Addressed
Interventions to Modify Health Care Provider Adherence to Asthma Guidelines (May 2013) ¹	Audit and feedback, education, organizational change
Primary Care Management of Abnormal Uterine Bleeding (March 2013) ²	Behavioral and lifestyle interventions--exercise, diet
Breathing Exercises and/or Retraining Techniques in the Treatment of Asthma: Comparative Effectiveness (September 2012) ³	Breathing techniques
Treatment for Depression After Unsatisfactory Response to SSRIs (April 2012) ⁴	CBT and other psychotherapies
Noncyclic Chronic Pelvic Pain Therapies for Women: Comparative Effectiveness (January 2012) ⁵	CBT, comprehensive therapy
Comparative Effectiveness of Psychological Treatments and Pharmacological Treatments for Adults with Posttraumatic Stress Disorder (April 2013) ⁶	CBT, exposure therapy, narrative therapy, cognitive processing, other psychotherapies, relaxation
Child Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Maltreatment (April 2013) ⁷	CBT, family therapy, psychotherapy, hypnosis
Interventions for the Prevention of Posttraumatic Stress Disorder (PTSD) in Adults After Exposure to Psychological Trauma (April 2013) ⁸	CBT, psychological interventions, hypnosis, education
Interventions for Adolescents and Young Adults with Autism Spectrum Disorders (August 2012) ⁹	CBT, social skills, job training
Screening, Behavioral Counseling, and Referral in Primary Care to Reduce Alcohol Misuse (July 2012) ¹⁰	Counseling, support
Screening and Diagnosing Gestational Diabetes Mellitus (November 2012) ¹¹	Diet modification
Comparative Effectiveness of Bariatric Surgery and non-Surgical Therapy in Adults with Metabolic Conditions and Body Mass Index of 30 to 34.9 kg/m ² (June 2013) ¹²	Diet, exercise, counseling, education
Recurrent Nephrolithiasis in Adults: Comparative Effectiveness of Preventive Medical Strategies (July 2012) ¹³	Dietary therapy
Comparative Effectiveness of Therapies for Children with Autism Spectrum Disorders (April 2011) ¹⁴	Early intervention, CBT, social skills, play and interaction-focused interventions
Critical Analysis of the Evidence for Patient Safety Practices (December 2011) ^{*15}	Education, audit and feedback
Closing the Quality Gap Series: Prevention of Health Care-Associated Infections (November 2012) ¹⁶	Education, audit and feedback, organizational change
Closing the Quality Gap Series: Medication Adherence Interventions: Comparative Effectiveness (September 2012) ¹⁷	Education, coaching, social support

Report	Examples of Behavioral Interventions Addressed
Long-Term Care for Older Adults: A Review of Home and Community-Based Services Versus Institutional Care (November 2012) ¹⁸	Education, counseling
Closing the Quality Gap Series: Medication Adherence Interventions: Comparative Effectiveness (September 2012) ^{*17}	Education, counseling, behavioral interventions, reminders, rewards, etc., variation in usual care,
Strategies to Prevent Weight Gain Among Adults (March 2013) ¹⁹	Education, exercise, diet
Strategies to Reduce Cesarean Birth in Low-Risk Women (October 2012) ²⁰	Education, relaxation, social support, doula, audit and feedback
Quality Improvement, Diabetes (September 2004) ^{*21}	Education, self-management
Comparison of Characteristics of Nursing Homes and Other Residential Long-Term Care Settings for People with Dementia (October 2012) ²²	Education, storytelling, group therapy, sensory stimulation, personalized care protocols
Literacy Interventions and Outcomes, Update (March 2011) ²³	Education, support
Treatment to Prevent Fractures in Men and Women with Low Bone Density or Osteoporosis: Update of a 2007 Report (March 2012) ²⁴	Exercise
Comparative Effectiveness of Treatments to Prevent Fracture in Men and Women with Low Bone Density or Osteoporosis Fractures (December 2007) ²⁵	Exercise
Treatment for Restless Leg Syndrome (November 2012) ²⁶	Exercise
Treatment Strategies for Patients with Peripheral Artery Disease (May 2013) ²⁷	Exercise
Exercise-induced Bronchoconstriction and Asthma (January 2010) ²⁸	Exercise warm up
Physical Therapy Interventions for Knee Pain Secondary to Osteoarthritis (November 2012) ²⁹	Exercise, tai chi
Interventions for Feeding and Nutrition in Cerebral Palsy (March 2013) ³⁰	Feeding interventions, positioning
Traumatic Brain Injury and Depression (April 2011) ³¹	Individual/group psychotherapy, neuropsychological rehabilitation, community-based rehabilitation, complementary and alternative medicine
Childhood Obesity Prevention Programs: Comparative Effectiveness Review and Meta-Analysis (June 2013) ³²	Lifestyle, weight management education
Multidisciplinary Postacute Rehabilitation for Moderate to Severe Traumatic Brain Injury in Adults (June 2012) ^{*33}	Multidisciplinary interventions
Health Information Technology Use for the Elderly, Chronically Ill, and Underserved (November 2008) ³⁴	Online support, health games, self-monitoring systems
Closing the Quality Gap Series: Quality Improvement Measurement of Outcomes for People with Disabilities (October 2012) ³⁵	Patient and provider education
Closing the Quality Gap Series: Quality Improvement Interventions to Address Health Disparities (August 2012) ³⁶	Patient and provider education; multicomponent interventions

Report	Examples of Behavioral Interventions Addressed
Transition of Care for Acute Stroke and Myocardial Infarction Patients from Hospitalization to Rehabilitation, Recovery and Secondary Prevention (October 2011) ³⁷	Patient education
Enhancing the Use and Quality of Colorectal Cancer Screening (February 2010) ³⁸	Patient education
Outcomes of Community Health Worker Interventions, Outcomes (June 2009) ³⁹	Patient education
Diabetes Education for Children with Type 1 Diabetes Mellitus and their Families (April 2008) ⁴⁰	Patient education, CBT, family therapy, motivational interviewing
Antibiotic Prescribing Behavior, Quality Improvement (January 2006) ⁴¹	Patient education, clinician education, audit and feedback
Closing the Quality Gap Series: Prevention of Healthcare-Associated Infections (November 2012) ¹⁶	Patient education, provider education, audit and feedback
Asthma Care, Quality Improvement (January 2007) ⁴³	Patient education, provider education, audit and feedback; brief discussion of heterogeneity, need for theoretical grounding for interventions
Nonsurgical Treatments for Urinary Incontinence in Adult Women: Diagnosis and Comparative Effectiveness (April 2012) ⁴⁴	Pelvic floor muscle training, vaginal cones, bladder training, weight loss
Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults (August 2011) ^{*45}	Physical therapy, weight loss
Comparative Effectiveness of Nonoperative and Operative Treatments for Rotator Cuff Tears (July 2010) ^{*46}	Postoperative rehabilitation, exercise, manual therapy, cortisone injections, acupuncture, other treatments and modalities typically delivered by physical therapists, osteopaths and chiropractors.
Practice-Based Interventions Addressing Concomitant Depression and Chronic Medical Conditions in the Primary Care Setting (August 2012) ⁴⁷	Problem solving therapy
Attention Deficit Hyperactivity Disorder: Effectiveness of Treatment in At-Risk Preschoolers; Long-Term Effectiveness in All Ages; and Variability in Prevalence, Diagnosis, and Treatment (October 2011) ⁴⁸	Psychotherapy
Child and Adolescent Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Trauma Other Than Maltreatment or Family Violence (February 2013) ⁴⁹	Psychotherapy, CBT, family therapy
Pain Management Interventions for Hip Fracture (May 2011) ⁵⁰	Rehabilitation, traction
Hypertension, Quality Improvement (January 2005) ^{*51}	Reminders, education, incentives
Self-Measured Blood Pressure Monitoring: Comparative Effectiveness (January 2012) ⁵²	Self-monitoring
Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults (September 2011) ^{*53}	Socio-environmental therapy, interpersonal psychotherapy, psychotherapy, cognitive therapy, CBT
Closing the Quality Gap Series: Improving Health Care and Palliative Care for Advanced and Serious Illness (October 2012) ⁵⁴	Support, counseling, communication

Report	Examples of Behavioral Interventions Addressed
Localized Prostate Cancer Surveillance (December 2011) ⁵⁵	Surveillance
Treatment for Overactive Bladder in Women (August 2009) ⁵⁶	Training, biofeedback
Interventions to Improve Cardiovascular Risk Factors in People with Serious Mental Illness (April 2013) ⁵⁷	Weight management, lifestyle interventions
Effectiveness of Weight Management Programs in Children and Adolescents (September 2008) ^{*58}	Weight management, lifestyle interventions

Note: CBT = Cognitive Behavioral Therapy

Appendix A References

1. Okelo SO, Butz AM, Sharma R, et al. Interventions To Modify Adherence to Asthma Guidelines. Rockville, MD: 2013 May. Report No.: AHRQ Publication No. 13-EHC022-EF Contract No.: 95. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
2. Hartmann KE, Jerome RN, Lindegren ML, et al. Primary care management of abnormal uterine bleeding. Rockville, MD: Agency for Healthcare Research and Quality, 2013 March. Report No.: No. 13-EHC025-EF. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
3. O'Connor E, Patnode CD, Burda BU, et al. Breathing exercises and/or retraining techniques in the treatment of asthma: Comparative effectiveness. Rockville, MD: Agency for Healthcare Research and Quality, 2012 September. Report No.: No. 12-EHC092-EF Contract No.: 71. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
4. Santaguida PL, MacQueen G, Keshavarz H, et al. Treatment for depression after unsatisfactory response to SSRIs. Rockville, MD: Agency for Healthcare Research and Quality, 2012 April. Report No.: No. 12-EHC050-EF Contract No.: 62. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
5. Andrews J, Yunker A, Reynolds WS, et al. Noncyclic chronic pelvic pain therapies for women: comparative effectiveness. Rockville, MD: Agency for Healthcare Research and Quality, 2012 January. Report No.: Contract No.: 41. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
6. Jonas DE, Cusack K, Forneris CA, et al. Psychological and pharmacological treatments for adults with posttraumatic stress disorder (PTSD). Rockville, MD: Agency for Healthcare Research and Quality, 2013 April. Report No.: No. 13-EHC011-EF Contract No.: 92. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
7. Fraser JG, Lloyd SW, Murphy RA, et al. Child exposure to trauma: comparative effectiveness of interventions addressing maltreatment. Rockville, MD: Agency for Healthcare Research and Quality, 2013 No. 13-EHC002-EF Contract No.: 89. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
8. Gartlehner G, Forneris CA, Brownley KA, et al. Interventions for the prevention of posttraumatic stress disorder (PTSD) in adults after exposure to psychological trauma. Rockville, MD: Agency for Healthcare Research and Quality, 2013 April. Report No.: Contract No.: 109. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
9. Taylor JL, Dove D, Veenstra-VanderWeele J, et al. Interventions for adolescents and young adults with autism spectrum disorders. Rockville, MD: 2012 August. Report No.: AHRQ Publication No. 12-EHC063-EF Contract No.: 65. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
10. Jonas DE, Garbutt JC, Brown JM, et al. Screening, behavioral counseling, and referral in primary care to reduce alcohol misuse. Rockville, MD: Agency for Healthcare Research and Quality, 2012 July. Report No.: No. 12-EHC055-EF Contract No.: 64. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
11. Hartling L, Dryden DM, Guthrie A, et al. Screening and diagnosing gestational diabetes mellitus. Rockville, MD: Agency for Healthcare Research and Quality, 2012 October. Report No.: No. 12(13)-E021-EF Contract No.: 210. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
12. Maglione MA, Gibbons MM, Livhits M, et al. Bariatric Surgery and Nonsurgical Therapy in Adults With Metabolic Conditions and a Body Mass Index of 30.0 to 34.9 kg/m². 2013 June. Report No.: AHRQ Publication No. 12(13)-EHC139-EF. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>

13. Fink HA, Wilt TJ, Eidman KE, et al. Recurrent nephrolithiasis in adults: comparative effectiveness of preventive medical strategies. Rockville, MD: Agency for Healthcare Research and Quality, 2012 July. Report No.: No. 12-EHC049-EF Contract No.: 61.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
14. Warren Z, Veenstra-VanderWeele J, Stone W, et al. Therapies for children with autism spectrum disorders. Rockville, MD: 2011 April. Report No.: AHRQ Publication No. 11-EHC029-EF Contract No.: 26.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
15. Critical analysis of the evidence for patient safety practices. Rockville, MD: Agency for Healthcare Research and Quality, 2011 November. Report No.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
16. Rothenberg BM, Marbella A, Pines E, et al. Prevention of Healthcare-Associated Infections. Closing the Quality Gap: Revisiting the State of the Science. Rockville, MD: 2012 AHRQ Publication No. 12(13)-E012-EF Contract No.: 208.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
17. Viswanathan M, Golin CE, Jones CD, et al. Medication Adherence Interventions: Comparative Effectiveness. 2012 September. Report No.: AHRQ Pub. No. 12-E010-1.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
18. Wysocki A, Butler M, Kane RL, et al. Long-term care for older adults: a review of home and community-based services versus institutional care. Rockville, MD: Agency for Healthcare Research and Quality, 2012 November. Report No.: No. 12(13)-EHC134-EF Contract No.: 81.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
19. Hutfless S, Maruthur NM, Wilson RF, et al. Strategies to Prevent Weight Gain Among Adults. Rockville, MD: 2013 March. Report No.: AHRQ Publication No. 13-EHC029-EF Contract No.: 97.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
20. Hartmann KE, Andrews JC, Jerome RN, et al. Strategies to reduce cesarean birth in low-risk women. Rockville, MD: Agency for Healthcare Research and Quality, 2012 October. Report No.: No. 12(13)-EHC128-EF Contract No.: 80.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
21. Coffey RM, Matthews TL, McDermott K. Diabetes care quality improvement: a resource guide for state action. Rockville, MD: Agency for Healthcare Research and Quality, 2004 September. Report No.: No. 04-0072.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
22. Zimmerman S, Anderson W, Brode S, et al. Comparison of characteristics of nursing homes and other residential long-term care settings for people with dementia. Rockville, MD: Agency for Healthcare Research and Quality, 2012 October. Report No.: No. 12(13)-EHC127-EF Contract No.: 79.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
23. Berkman ND, Sheridan SL, Donahue KE, et al. Health literacy and interventions and outcomes: an updated systematic review. Rockville, MD: Agency for Healthcare Research and Quality, 2011 March. Report No.: No. 11-E006 Contract No.: 199.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
24. Crandall CJ, Newberry SJ, Diamant A, et al. Treatment to prevent fractures in men and women with low bone density or osteoporosis: update of a 2007 report. Rockville, MD: Agency for Healthcare Research and Quality, 2012 March. Report No.: No. 12-EHC023-EF Contract No.: 53.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
25. MacLean C, Alexander A, Carter J, et al. Comparative effectiveness of treatments to prevent fractures in men and women with low bone density. Rockville, MD: Agency for Healthcare Research and Quality, 2007 December. Report No.: No. 08-EHC008-EF Contract No.: 12.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>

26. Wilt TJ, MacDonald R, Ouellette J, et al. Treatment for restless legs syndrome. Rockville, MD: Agency for Healthcare Research and Quality, 2012 November. Report No.: No. 12(13)-EHC147-EF Contract No.: 86.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
27. Jones WS, Schmit KM, Vemulapalli S, et al. Treatment strategies for patients with peripheral artery disease. Rockville, MD: Agency for Healthcare Research and Quality, 2013 May. Report No.: No. 13-EHC090-EF Contract No.: 118.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
28. Dryden DM, Spooner CH, Stickland MK, et al. Exercise-induce bronchoconstriction and asthma. Rockville, MD: Agency for Healthcare Research and Quality, 2010 January. Report No.: No. 10-E001 Contract No.: 189.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
29. Shamliyan TA, Wang S-Y, Olson-Kellogg B, et al. Physical therapy interventions for knee pain secondary to osteoarthritis. Rockville, MD: Agency for Healthcare Research and Quality, 2012 November. Report No.: No. 12(13)-EHC115-EF Contract No.: 77.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
30. Ferluga ED, Archer KR, Sathe NA, et al. Interventions for Feeding and Nutrition in Cerebral Palsy. Rockville, MD: 2013 March. Report No.: AHRQ Publication No. 13-EHC015-EF Contract No.: 94.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
31. Guillamondegui OD, Montgomery SA, Phibbs FT, et al. Traumatic brain injury and depression. Rockville, MD: Agency for Healthcare Research and Quality, 2011 April. Report No.: No. 11-EHC017-EF Contract No.: 25.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
32. Wang Y, Wu Y, Wilson RF, et al. Childhood obesity prevention programs: comparative effectiveness review and meta-analysis. Rockville, MD: 2013 June. Report No.: AHRQ Publication No. 13-EHC081-EF Contract No.: 115.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
33. Brasure M, Lamberty GJ, Sayer NA, et al. Multidisciplinary postacute rehabilitation for moderate to severe traumatic brain injury in adults. Rockville, MD: Agency for Healthcare Research and Quality, 2012 June. Report No.: No. 12-EHC101-EF Contract No.: 72.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
34. Jimison H, Gorman P, Woods S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Rockville, MD: Agency for Healthcare Research and Quality, 2008 November. Report No.: No. 09-E004 Contract No.: 175.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
35. Butler M, Kane RL, Larson S, et al. Quality improvement measurement of outcomes for people with disabilities. Closing the Quality Gap: Revisiting the State of the Science. Rockville, MD: Agency for Healthcare Research and Quality, 2012 October. Report No.: No. 12(13)-E013-EF Contract No.: 208.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
36. McPheeters ML, Kripalani S, Peterson NB, et al. Quality improvement interventions to address health disparities. Closing the Quality Gap: Revisiting the State of the Science. Rockville, MD: Agency for Healthcare Research and Quality, 2012 August. Report No.: No. 12-E009-EF Contract No.: 208.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>

37. Olson DM, Bettger JP, Alexander KP, et al. Transition of care for acute stroke and myocardial infarction patients. Rockville, MD: Agency for Healthcare Research and Quality, 2011 October. Report No.: No. 11(12)-E011 Contract No.: 202. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
38. Holden DJ, Harris R, Porterfield DS, et al. Enhancing the use and quality of colorectal cancer screening. Rockville, MD: Agency for Healthcare Research and Quality, 2010 February. Report No.: No. 10-E002 Contract No.: 190. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
39. Viswanathan M, Kraschnewski J, Nishikawa B, et al. Outcomes of community health worker interventions. Rockville, MD: Agency for Healthcare Research and Quality, 2009 June. Report No.: No. 09-E014 Contract No.: 181. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
40. Couch R, Jetha M, Dryden DM, et al. Diabetes education for children with Type 1 Diabetes Mellitus and their families. Rockville, MD: Agency for Healthcare Research and Quality, 2008 April. Report No.: No. 08-E011 Contract No.: 166. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
41. Ranji SR, Steinman MA, Shojanian KG, et al. Antibiotic Prescribing Behavior. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Rockville, MD: Agency for Healthcare Research and Quality, 2006 January. Report No.: No. 04(06)-0051-4 Contract No.: 9.4. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
42. Gibbons MC, Wilson RF, Samal L, et al. Impact of consumer health informatics applications. Rockville, MD: Agency for Healthcare Research and Quality, 2009 October. Report No.: No. 09(10)E-019 Contract No.: 188. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
43. Bravata DM, Sundaram V, Lewis R, et al. Asthma Care. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Rockville, MD: Agency for Healthcare Research and Quality, 2007 January. Report No.: No. 04(07)-0051-5 Contract No.: 9.5. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
44. Shamliyan T, Wyman J, Kane RL. Nonsurgical treatments for urinary incontinence in adult women: diagnosis and comparative effectiveness. Rockville, MD: Agency for Healthcare Research and Quality, 2012 April. Report No.: No. 11(12)-EHC074-EF Contract No.: 36. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
45. Balk EM, Moorthy D, Obadan NO, et al. Diagnosis and treatment of obstructive sleep apnea in adults. Rockville, MD: Agency for Healthcare Research and Quality, 2011 July. Report No.: No. 11-EHC052 Contract No.: 32. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
46. Seida JC, Schouten JR, Mousavi SS, et al. Comparative effectiveness of nonoperative and operative treatments for rotator cuff tears. Rockville, MD: Agency for Healthcare Research and Quality, 2010 July. Report No.: No. 10-EHC050-EF Contract No.: 22. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
47. Watson L, Amick HR, Gaynes BA, et al. Practice-based interventions addressing concomitant depression and chronic medical conditions in the primary care settings. Rockville, MD: Agency for Healthcare Research and Quality, 2012 August. Report No.: No. 12-EHC106-EF Contract No.: 75. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>

48. Charach A, Dashti B, Carson P, et al. Attention Deficit Hyperactivity Disorder: Effectiveness of Treatment in At-risk Preschoolers; Long-term Effectiveness in All Ages; and Variability in Prevalence, Diagnosis, and Treatment. Rockville, MD: 2011 October. Report No.: AHRQ Publication No. 12-EHC003-EF Contract No.: 44. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
49. Forman-Hoffman V, Knauer S, McKeeman J, et al. Child and Adolescent Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Trauma Other Than Maltreatment or Family Violence. Rockville, MD: 2013 February. Report No.: AHRQ Publication No. 13-EHC054-EF Contract No.: 107. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
50. Abou-Setta AM, Beaupre LA, Jones CA, et al. Pain management interventions for hip fracture. Rockville, MD: Agency for Healthcare Research and Quality, 2011 May. Report No.: No. 11-EHC022-EF Contract No.: 30. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
51. Walsh J, McDonald KM, Shojanika KG, et al. Hypertension Care. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Rockville, MD: Agency for Healthcare Research and Quality, 2005 January. Report No.: No. 04-0051-3 Contract No.: 9.3. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
52. Uhlig K, Balk EM, Patel K, et al. Self-measured blood pressure monitoring: comparative effectiveness. Rockville, MD: Agency for Healthcare Research and Quality, 2012 January. Report No.: No. 12-EHC002-EF Contract No.: 45. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
53. Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults. Houston, TX: John M. Eisenberg Center for Clinical Decision and Communication Science, Baylor College of Medicine, 2011 September. Report No.: AHRQ Pub. No. 11(12)-EHC056-3 Contract No.: 33. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
54. Improving health care and palliative care for advanced and serious illness. Closing the Quality Gap: Revisiting the State of the Science. Rockville, MD: Agency for Healthcare Research and Quality, 2012 October. Report No.: No. 12(13)-E014-1 Contract No.: 208. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
55. Ip S, Dahabreh IJ, Chung M, et al. An evidence review of active surveillance in men with localized prostate cancer. Rockville, MD: Agency for Healthcare Research and Quality, 2011 December. Report No.: No. 12-E003-EF Contract No.: 204. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
56. Hartmann KE, McPheeters ML, Biller DH, et al. Treatment for overactive bladder in women. Rockville, MD: 2009 August. Report No.: AHRQ Publication No. 09-E017 Contract No.: 187. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
57. Gierisch JM, Nieuwsma JA, Bradford DW, et al. Interventions to improve cardiovascular risk factors in people with serious mental illness. Rockville, MD: 2013 April. Report No.: AHRQ Publication No. 13-EHC063-EF Contract No.: 105. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>
58. Whitlock EP, O'Connor EA, Williams SB, et al. Effectiveness of weight management programs in children and adolescents. Rockville, MD: 2008 September. Report No.: AHRQ Publication No. 08-E014 Contract No.: 170. <http://effectivehealthcare.ahrq.gov/reports/final.cfm>

Appendix B. Approaches to Addressing Behavior Change Interventions in EPC Reviews

EPC reports that included a behavior change intervention focused on a broad variety of topics. These included mental health (depression), behavioral condition, or a medical condition (such as asthma, nephrolithiasis, and weight gain prevention) or a nonclinical topic that included an intervention aimed at behavior change (such as consumer health informatics, public reporting, and provider adherence to asthma guidelines). From these reviews we provide examples of various approaches to challenges.

Review Scope and Key Questions

Many reviews included additional questions to provide relevant information for stakeholder decisionmaking, in addition to questions related to effectiveness. These questions explored issues such as contextual factors;¹⁻³ analysis of intervention features of more effective interventions;⁴⁻⁵ usability (public reporting);⁶ and implementation strategies.⁷ For example, a review on autism⁵ included a question on treatment modifiers in addition to a question on intervention effectiveness. Treatment modifiers included intervention features, context and patient characteristics. Investigators noted that few studies were designed or powered to identify modifiers of treatment effect. They described the three studies that were, and further described potential factors described in the other included studies that warranted further study.

In a review of autism in children,⁵ in addition to effectiveness questions, the authors assessed the evidence for the presumed mechanism of action, linking the intervention targeting the behavior (such as maladaptive behavior), functional outcome (psychosocial adaptation), and the long-term outcomes. In their analytic framework, they assumed that improvements in treatment outcomes would lead to improved functional outcomes, and then long-term clinical outcomes. They included questions about predictors of favorable treatment outcomes (shorter term intermediate behavioral outcomes), and whether favorable treatment outcome predicted favorable functional outcomes. These questions explored not only predictors of response, but potentially the assumed association between behavioral and functional outcomes.

Study Selection

A design other than the RCT may have been chosen by a primary researcher because of feasibility, ethical issues, or cost. This poses difficulties for a review focused on establishing causality. Almost all reviews included key questions focused on global effectiveness, and a variety of study designs were included to answer this question, most commonly RCTs. Most often the rationale if given for inclusion of a greater variety of study designs was the relative lack of RCTs.⁸ How far investigators chose to go in the evidence hierarchy was a balance between the volume of literature and presumed risk of bias (and ability to establish causality) based on study design. For example in an autism report⁵ single-subject design studies were excluded. They noted that these studies could assess treatment response in short timeframes and under very tightly controlled circumstances, but could not provide information on longer term or functional outcomes, or external validity.

Investigators used a different body of studies depending on the type of questions in the review's scope. For example, in an evidence review on health information technology use,⁹ the authors acknowledged that RCTs were optimal for assessing outcomes, but other study designs

were appropriate for other types of questions. Thus they included epidemiologic studies and qualitative research to inform questions about barriers, drivers, and usability. In a review of the PCMH, investigators included studies without a comparator when describing implementation strategies.³

Meaningfully Categorizing and Describing Interventions

EPC reviews have categorized interventions, both with and without behavior change interventions, by programs;¹⁰ individual components¹¹ or bundles of components² using existing taxonomies. Choice of approach depended on the granularity of the review question and end-user need. For example in a review of screening for alcohol misuse,¹² behavioral interventions were grouped as low and high intensity. This was done because the review was used to inform a recommendation on screening, rather than effectiveness of one behavioral counseling approach over another. Another review, on modification of cardiovascular risk factors in individuals with SMI, Gierisch et al.¹³ took a more general approach and analyzed all behavioral interventions together as a group, and compared them to usual care.

Grouping multi-component interventions was a challenge. In another review on medication adherence for hepatitis C treatment, interventions were grouped by intervention target-policy, regimen, patient, adverse event management, or provider.¹⁴ Investigators were unable to group at a more granular level for most categories, such as intervention bundle, component, or level of intensity, because no two studies utilized the same intervention. Because of limited descriptions and differences in approach, they were unable to identify the most effective intervention components.

In a review on prevention of healthcare-associated infection,² investigators identified 16 different combinations of components across studies. Sufficient similarities were found to allow investigators to group strategies by their components. They designated organizational change and provider education as base strategies because they were found in a majority of intervention bundles.

A review on medication adherence⁴ categorized interventions by a previously developed taxonomy, and noted those that were tailored. The presence or absence of tailoring was not used to stratify the analysis, likely related to the fact that they were often single studies of a diversity of interventions. As a result in the analysis of features of effective interventions, the effect of tailoring could not be explored. This review, however, did describe characteristics of interventions, related to the target, mode of delivery, intensity, duration and components. However, because of the diversity of interventions and limited descriptions of both the intervention and characteristics, investigators were unable to compare these features and make inferences about their potential impact on outcomes. While recognizing the potential influences of these intervention characteristics, they were unable to do more than describe what was reported in the primary literature.

Outcome Selection and Measurement

Reviews included behavioral outcomes in reviews as intermediate outcomes. For example in a review of treatments for disruptive behaviors,⁸ behavioral outcomes such as aggressive behavior, fighting and property destruction, were proximal to the functional outcomes of family functioning, school performance, and quality of life. In a review of smoking cessation interventions included intermediate outcomes related to behavior change:¹⁵ cessation, relapse,

and continuous abstinence, in addition to clinical outcomes such as asthma, hospitalization, and pre-term birth.

Some reviews required reporting of an objective clinical outcome in addition to the behavioral outcome for inclusion into the body of evidence. Presumably this was done to ensure that the more distal outcomes could be attributed to the impact of the intervention on the desired behavior change. For example in a review on diabetes strategies,¹⁶ investigators excluded studies that did not include both outcomes of patient adherence, provider/patient understanding, satisfaction, or self-efficacy, and a measure of disease control or adherence. They also excluded all self-reported adherence measures for providers. In another review on adult weight gain prevention,¹³ investigators had planned to include the intermediate outcomes of knowledge, attitude, and skills when a weight outcome was also reported.

Quality Assessment of Individual Studies

To assess individual study quality, authors drew from existing resources including the EPC program methods guidance, USPSTF methods, Cochrane methods guidance, Newcastle-Ottawa Scale, and CRD. Most times qualitative research included to inform descriptive questions such as issues as barriers, usability, and context, were not formally assessed.^{2,9}

Authors adapted available quality assessment tools and drew from previously adapted tools for the program using varied approaches. In the HAI prevention report,² authors adapted a quality assessment tool from a previous EPC review on HAI prevention and included additional criteria were drawn from the RTI Item Bank for assessing observational studies. Investigators developed a new tool and used it across a series of autism reviews.^{5,17} When authors create new tools or revisions to new tools, one-off approaches may lack consistent interpretations and validity.

Synthesis

Many reviews that had a diversity of study designs conducted a qualitative synthesis. In select reviews, lower quality studies were excluded from the body of evidence.^{4,18}

In a review of smoking cessation¹⁵ in pregnancy and post-partum, investigators were able to assess the impact of individual intervention components. They used a logistic mixed effects model that estimated quit rate across studies. They were able to evaluate some components individually (such as feedback information and peer support) but could not assess counseling as it was commonly provided across intervention and control arms.

In a review on patient safety practices,³ investigators examined the evidence for implementation and cost, and qualitatively synthesized for themes about effective implementation. They categorized implementation difficulty as difficult, not difficult, and moderate; and cost as low, medium, and high. These elements, in addition to assessment of effectiveness, were included to help with decisionmaking.

SOE

Most reviews graded the strength of evidence using EPC methods guidance. SOE was generally not graded for descriptive key questions, such as description of intervention components and implementation features.^{7,19}

Appendix B References

1. Hussey P, Mulcahy A, Schnyer C, et al. Bundled Payment: Effects on Health Care Spending and Quality. AHRQ Publication No.12-E007-EF. Rockville, MD:Agency for Healthcare Research and Quality;August 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
2. Rothenberg BM, Marbella A, Pines E, et al. Prevention of Healthcare-Associated Infections. Closing the Quality Gap: Revisiting the State of the Science. AHRQ Publication No.12(13)-E012-EF. Rockville, MD:Agency for Healthcare Research and Quality; 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
3. Shekelle PG, Wachter RM, Pronovost PJ, et al. Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices. AHRQ Publication No.13-E001-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2013.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
4. Viswanathan M, Golin CE, Jones CD, et al. Medication Adherence Interventions: Comparative Effectiveness. AHRQ Pub. No.12-E010-1. Agency for Healthcare Research and Quality;September 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
5. Warren Z, Veenstra-VanderWeele J, Stone W, et al. Therapies for children with autism spectrum disorders. AHRQ Publication No.11-EHC029-EF. Rockville, MD: Agency for Healthcare Research and Quality;April 2011.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
6. Totten AM, Wagner J, Tiwari A, et al. Public reporting as a quality improvement strategy. Closing the quality gap: revisiting the state of the science. Evidence Report/Technology Assessment. 2012(208).
7. Williams J, Jackson G, Powers B, et al. The Patient-Centered Medical Home. AHRQ Publication No.12-E008-1. Rockville, MD: July 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
8. Psychosocial and pharmacologic interventions for disruptive behavior disorder. Rockville, MD: Agency for Healthcare Research and Quality; January 2014.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
9. Jimison H, Gorman P, Woods S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. AHRQ Publication No.09-E004. Rockville, MD: Agency for Healthcare Research and Quality; November 2008.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
10. Forman-Hoffman V, Knauer S, McKeeman J, et al. Child and Adolescent Exposure to Trauma: Comparative Effectiveness of Interventions Addressing Trauma Other Than Maltreatment or Family Violence. AHRQ Publication No.13-EHC054-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2013.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
11. Bravata DM, Sundaram V, Lewis R, et al. A Critical Analysis of Quality Improvement Strategies (Vol. 5: Asthma Care). AHRQ Publication No. 04(07)-0051-5. Rockville, MD: Agency for Healthcare Research and Quality; January 2007.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
12. Screening and Behavioral Counseling Interventions in Primary Care to Reduce Alcohol Misuse. U.S. Preventive Services Task Force; 2012.
<http://www.uspreventiveservicestaskforce.org/uspstf12/alc misuse/alc misusefinalrs.htm>. Accessed on August 2014.

13. Gierisch JM, Nieuwsma JA, Bradford DW, et al. Interventions to improve cardiovascular risk factors in people with serious mental illness. AHRQ Publication No.13-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; April 2013.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
14. Sun X, Patnode CD, Williams C, et al. Interventions to Improve Patient Adherence to Hepatitis C Treatment: Comparative Effectiveness. AHRQ Publication No.13-EHC009-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
15. Likis FE, Andrews JC, Fonnesebeck CJ, et al. Smoking Cessation Interventions in Pregnancy and Postpartum Care. AHRQ Publication No.14-E001-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
16. Shojania KG, Ranji SR, Shaw LK, et al. A Critical Analysis of Quality Improvement Strategies (Vol. 2: Diabetes Care). Rockville, MD: Agency for Healthcare Research and Quality; September 2004.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
17. Taylor JL, Dove D, Veenstra-VanderWeele J, et al. Interventions for adolescents and young adults with autism spectrum disorders. AHRQ Publication No.12-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; August 2012.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
18. Hutfless S, Maruthur NM, Wilson RF, et al. Strategies to Prevent Weight Gain Among Adults. AHRQ Publication No.13-EHC029-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2013.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>
19. Medication Therapy Management. Rockville, MD: Agency for Healthcare Research and Quality; July 2013.
<http://effectivehealthcare.ahrq.gov/reports/final.cfm>

Appendix C. Key Informants

Andrew Booth, Dip. Lib., M.Sc., MCLIP
Co-Covenor
International Cochrane Collaboration
Qualitative
Research Methods Group
Sheffield, UK

Robert Hahn, Ph.D., M.D.
Senior Scientist
U.S. Task Force on Community Preventive
Services
Atlanta, GA

Mark Petticrew, Ph.D.
Director
Public Health Research Consortium
London, UK

Therese Pigott, Ph.D.
Co-Chair and Editor
Campbell Collaboration Methods Group
Loyola University Chicago
Chicago, IL

Brian Reichow, Ph.D., BCBA-D
Assistant Professor (Adjunct), Associate
Research Scientist
Child Study Center, Yale School of
Medicine
New Haven, CT

Ken Renisow, Ph.D.
Chair
Society of Behavioral Medicine, Evidence-
based Behavioral Medicine (EBBM)
Committee
Ann Arbor, MI

Christopher Schmid, Ph.D.
Director, Professor
Biostatistics, Brown University
Center for Evidence Based Medicine
Providence, RI

Appendix D. Key Informant Interview Guide

Introduction

The overall mission of the Agency for Healthcare Research and Quality's (AHRQ) Effective Health Care (EHC) Program is to provide evidence-based information to health care stakeholders that is relevant to their needs, timely, objective, scientifically rigorous in construct, and developed and presented with transparency. In the production of systematic reviews, we aim to answer questions about effectiveness of interventions and average population effects. We are aware that for certain conditions and behavioral interventions, these questions may miss important issues.

AHRQ engages stakeholders in all facets of their research enterprise, including the producing of systematic reviews, with the goals of ensuring that research findings reflect the needs of diverse users, are relevant to their unique challenges, and are applicable in real-world situations.

Purpose of the discussion session

The goal of our project is to better understand the questions that should be asked about behavioral interventions. Additionally, we would like to understand methods used and challenges encountered in including behavioral literature in systematic reviews, including which study design types are included in reviews.

We are very interested in learning from your experience.

There are not right or wrong answers, so please feel free to share your thoughts openly.

We would welcome any materials that you would like to share with us either before or after the discussion session. Please send any materials to Johanna.anderson2@va.gov.

Ground rules for discussion session

The discussions will be tape recorded, transcribed, and analyzed for overarching themes.

Although the report may list individuals who were interviewed, answers will not be identifiable to individuals or specific organizations.

You may refrain from answering any questions and are welcome to leave the discussion at any time.

Interview Guide

1. How long have you been conducting systematic reviews which include behavioral interventions?
2. What type of behavioral interventions have you studied? For what conditions?
3. In systematic reviews, what are the right questions to ask for patients who receive behavioral interventions?
 - a. What works?
 - b. For whom and how?
 - c. How separable or combinable are these questions for patients who receive behavioral interventions?
4. What research study designs flow from these questions in your evidence synthesis process?
5. At what point do we shift attention from the population to the individual? How do we present effectiveness as a distribution rather than a point estimate?
6. What kind of data could answer these questions and how might that data be analyzed or presented? (Example?)
7. What are the implications for assessing strength of evidence? How do you pull together qualities to understand how strong evidence is?
8. What are your thoughts on how to assess internal validity in these studies?
9. In our process, we separate applicability and internal validity. When do you feel it is okay to separate external from internal validity?
10. What is your viewpoint on using behavioral interventions to establish an effect in a systematic review?