Methods Research Report

Comparative Effectiveness Review Methods: Clinical Heterogeneity

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. 290-2007-10056-I

Prepared by:
RTI International–University of North Carolina Evidence-based Practice Center
Research Triangle Park, North Carolina

Investigators:
Suzanne L. West, M.P.H., Ph.D.
Gerald Gartlehner, M.D., M.P.H.
Alyssa J. Mansfield, M.P.H., Ph.D.
Charles Poole, M.P.H., Sc.D.
Elizabeth Tant, B.S.
Nancy Lenfestey, M.H.A.
Linda J. Lux, M.P.A.
Jacqueline Amoozegar, B.A.
Sally C. Morton, Ph.D.
Timothy C. Carey, M.D., M.P.H.
Meera Viswanathan, Ph.D.
Kathleen N. Lohr, Ph.D.

AHRQ Publication No. 10-EHC070-EF
September 2010

This report is based on research conducted by the RTI International–University of North Carolina Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10056-I). The findings and conclusions in this document are those of the author(s), 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 clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

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

This document was written with support from the Effective Health Care Program at AHRQ. None of the authors has a financial interest in any of the products discussed in this document.

The investigators have no relevant financial interests in the report. The investigators have no employment, consultancies, honoraria, or stock ownership or options, or royalties from any organization or entity with a financial interest or financial conflict with the subject matter discussed in the report.

Suggested citation: West SL, Gartlehner G, Mansfield AJ, Poole C, Tant E, Lenfestey N, Lux LJ, Amoozegar J, Morton SC, Carey TC, Viswanathan M, Lohr KN. Comparative Effectiveness Review Methods: Clinical Heterogeneity. Agency for Healthcare Research and Quality; September 2010. Methods Research Paper. AHRQ Publication No. 10-EHC070-EF. Available at http://effectivehealthcare.ahrq.gov/.

Preface

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

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

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

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

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

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

Stephanie Chang, M.D., M.P.H.
EPC Program Director and Task Order Officer
Agency for Healthcare Research and Quality

Comparative Effectiveness Review Methods: Clinical Heterogeneity

Structured Abstract

Objectives. The Agency for Healthcare Research and Quality (AHRQ) funded the RTI International—University of North Carolina at Chapel Hill Evidence-based Practice Center to determine best practices for addressing clinical heterogeneity in systematic reviews (SRs) and comparative effectiveness reviews (CERs). These best practices address critiques from patients, clinicians, policymakers, and others who assert that SRs typically focus on broad populations and, as a result, often lack information relevant to individual patients or patient subgroups.

Data sources and methods. We used numerous data sources. We abstracted information from guidance documents prepared by U.S. and international organizations engaged in preparing reviews. We searched MEDLINE® to identify studies on how to handle clinical heterogeneity and subgroup analyses. We reviewed more than 120 SRs conducted by AHRQ's Evidence-based Practice Centers (EPCs), the Cochrane Collaboration, the Drug Effectiveness Review Project, the United Kingdom's National Institute for Health and Clinical Excellenceand others that we identified from the Centre for Reviews and Dissemination Database of Abstracts of Reviews of Effects and Health Technology Assessment. We reviewed peer and public review comments from AHRQ's Scientific Review Center for three CERs, and we conducted key informant interviews with authors of six SRs prepared by AHRQ's EPCs or international organizations.

Results. Clinical heterogeneity has been defined as the variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect (e.g., odds ratio, risk ratio, risk difference). Statistical heterogeneity is defined as variability in the observed treatment effects beyond what would be expected by random error. The review organizations we studied varied in their inclusion of factors, in terms of the key questions and analysis that may modify the treatment-outcome association. They tended to give more consideration to demographic factors than to disease factors (e.g., disease severity, risk factors, coexisting disease, or cointerventions). Individual systematic reviewers whom we interviewed preferred a priori identification of effect modifiers to post hoc determination because of the latter's data-dredging nature and the possibility of type 1 error when many subgroups are evaluated. Many publications that we identified through our literature searches did indicate that analysis of individual patient-level data in meta-analyses does allow better assessment of clinical heterogeneity, but the time, cost, and difficulty in obtaining these data are often prohibitive.

Conclusions. Identifying factors that may influence the treatment-outcome association is important to clinicians and patients because it helps them understand which patients will benefit most, who is least likely to benefit, and who is at greatest risk of experiencing adverse outcomes. Clear evidence-based guidance on addressing clinical heterogeneity in SRs and CERs is not available currently but would be valuable to AHRQ's EPCs and to others conducting SRs internationally.

Contents

Tables

Figures

Appendixes

Executive Summary

Introduction

The Agency for Healthcare Research and Quality (AHRQ) commissioned the RTI International–University of North Carolina at Chapel Hill (RTI-UNC) Evidence-based Practice Center (EPC) to explore how systematic review groups have dealt with clinical heterogeneity and to seek out best practices for addressing clinical heterogeneity in systematic reviews (SRs) and comparative effectiveness reviews (CERs). Such best practices, to the extent they exist, may enable AHRQ's EPCs to address critiques from patients, clinicians, policymakers, and other proponents of health care about the extent to which "average" estimates of the benefits and harms of health care interventions apply to individual patients or to small groups of patients sharing similar characteristics.

Such users of reviews often assert that EPC reviews typically focus on broad populations and, as a result, often lack information relevant to patient subgroups that are of particular concern to them. More important, even when EPCs evaluate literature on homogeneous groups, there may be varying individual treatment for no apparent reason, indicating that average treatment effect does not point to the *best* treatment for any given individual. Thus, the health care community is looking for better ways to develop information that may foster better medical care at a "personal" or "individual" level. (We do not use the phrases "personalized medicine" or "individualized medicine" here, because of the terms' commonly understood applications in genetics and genomics.)

To address our charge for this methods project, the EPC set out to answer six key questions (KQ) (Table A). AHRQ assigned these KQs to us and we worked with AHRQ staff and the EPC program's Scientific Resource Center (SRC) at the Oregon Health & Science University on approaches to address the five empirical issues and their subquestions. KQ 6 asked the project team to put forward ideas that an AHRQ cross-EPC work group might take on in 2010 or later, drawing on our findings for the first five questions. As implied by KQ 6, AHRQ wanted to understand how its EPC program (and the EPC tasks in the Effective Health Care Program related to production of CERs) can better address concerns of stakeholders related to clinical heterogeneity—i.e., how confidently clinicians, policymakers, and others can draw conclusions about the effectiveness of interventions from reports that account for clinical heterogeneity in both the populations of interest to them and the populations studied. Although the first set of audiences are oriented to the United States, we believe that our findings can be helpful to systematic reviewers globally.

Table A. Key questions for methods report on clinical heterogeneity

1. What is clinical heterogeneity?
   a. How has it been defined by various groups?
   b. How is it distinct from statistical heterogeneity?
   c. How does it fit with other issues that have been addressed by the AHRQ Methods Manual for CERs?
2. How have systematic reviews dealt with clinical heterogeneity in the key questions?
   a. What questions have been asked?
   b. How have they pre-identified population subgroups with common clinical characteristics that modify their intervention-outcome association?
   c. What are best practices in key questions and how these subgroups have been identified?
3. How have systematic reviews dealt with clinical heterogeneity in the review process?
   a. What do guidance documents of various systematic review groups recommend?
   b. How have EPCs handled clinical heterogeneity in their reviews?
   c. What are best practices in searching for and interpreting results for particular subgroups with common clinical characteristics that may modify their intervention-outcome association?
4. What are critiques in how systematic reviews handle clinical heterogeneity?
   a. What are critiques from specific reviews (peer and public) on how EPCs handled clinical heterogeneity?
   b. What general critiques (in the literature) have been made against how systematic reviews handle clinical heterogeneity?
5. What evidence is there to support how to best address clinical heterogeneity in a systematic review?
6. What questions should an EPC work group on clinical heterogeneity address?

Before focusing on clinical heterogeneity per se, we needed to clarify three other terms often appearing in EPC reviews: effect measure, methodologic heterogeneity, and statistical heterogeneity. Table B provides definitions of these concepts and, specifically, gives the definition we used for clinical heterogeneity for this project.

Table B. Core concepts of heterogeneity and their definitions

Clinical heterogeneity	Variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect.
Measure of effect	A value that measures the effect of a variable on the frequency or risk of a health outcome, such as an odds ratio, risk ratio, risk difference, or absolute difference.
Methodologic heterogeneity	In the context of EPC reviews, among-study differences in estimated effect sizes for the intervention that can be attributed to variability and quality of study designs and analyses.
Statistical heterogeneity	Variability in the observed treatment effects beyond what would be expected by random error.

Heterogeneity (of any type) in EPC reviews is important because its appearance suggests that included studies differed on one or more dimensions such as patient demographics, study designs, coexisting conditions, or other factors. EPCs then need to clarify for clinical and other audiences, collectively referred to as stakeholders, what are the potential causes of the heterogeneity in their results. This will allow the stakeholders to understand whether and to what degree they can apply this information to their own patients or constituents. Of greatest importance for this project was clinical heterogeneity, which we define as the variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across

studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect (e.g., odds ratio, risk ratio, risk difference)Assessing how systematic reviewers approach clinical heterogeneity required us to develop and adopt a working definition of clinical heterogeneity and to explore how reviewers typically treat various types of heterogeneity. One major issue in dealing with heterogeneity, and clinical heterogeneity in particular, was that these terms have not been used consistently in clinical research or the SR literature. In fact, the term "clinical heterogeneity" may be more appropriate when used in the context of individual clinical studies rather than for SRs and CERs. Some review groups, such as the Cochrane Collaboration and the Centre for Reviews and Dissemination, use "clinical diversity" rather than "clinical heterogeneity" to describe clinical differences among studies in SRs. Because we could not find a clear definition of clinical heterogeneity in guidance documents or in the published literature that distinguished clinical heterogeneity from clinical diversity, we treat "clinical heterogeneity" and "clinical diversity" as synonymous in this report.

Researchers often consider any heterogeneity problematic because it indicates that pooling across studies may not be appropriate, yet true heterogeneity can be informative by suggesting new avenues for research investigations. Ideally, one could differentiate between heterogeneity of treatment effects stemming from factors, such as demographics, coexisting conditions, treatments, or genetics (what we and other researchers term "clinical heterogeneity") and that resulting from variability in study design and analysis (which we and others refer to as "methodologic heterogeneity"). However, trying to distinguish between clinical and methodological heterogeneity is not easy because they are intertwined; both can and do co-occur in SRs and CERs.

Alternatively, "statistical heterogeneity" refers to the variability in observed treatment effects that is beyond what would be expected by random error (chance). Statistical heterogeneity may *signal* the presence of clinical heterogeneity, methodological heterogeneity, or chance. The difference between clinical heterogeneity and statistical heterogeneity can be thought of as a cause and an effect relationship, respectively: when clinical heterogeneity is apparent across studies included in a meta-analysis, it can lead to some degree of statistical heterogeneity.

How to address clinical heterogeneity when conducting SRs, CERs, and meta-analyses has been discussed in the literature, but little consensus has been reached on best practices for identifying and understanding the factors underlying such heterogeneity, which was one of the goals of this project.

Methods

To produce this methods report on issues relating to clinical heterogeneity and the six key questions, we used a variety of data sources. For KQ 1, we reviewed guidance documents developed by organizations involved in developing SRs and clinical practice guidelines. These organizations included AHRQ, Centre for Reviews and Dissemination (CRD), Cochrane Collaboration, Drug Effectiveness Review Program (DERP), Institute for Quality and Efficiency in Health Care (IQWIG), National Health and Medical Research Council (NHMRC), UK National Institute for Health and Clinical Excellence (NICE), and various health technology assessment organizations.

The literature base for KQs 2 and 3 included selected, relevant literature—i.e., SRs and CERs and similar reports (e.g., health technology assessments)—completed by four organizations with extensive expertise in literature syntheses: AHRQ, Cochrane Collaboration,

DERP, and NICE. Our sample also included syntheses catalogued in the Centre for Reviews and Dissemination's Database of Abstracts of Reviews of Effects (DARE) and the Health Technology Assessment (HTA) database. We limited our evaluation to reviews of 15 clinical conditions: breast cancer, lung cancer, prostate cancer, congestive heart failure, cesarean section, chronic kidney disease, chronic obstructive pulmonary disease (COPD), depression, dyspepsia/gastroesophageal reflux disease (GERD), heavy menstrual bleeding, hypertension, irritable bowel syndrome, labor induction, myocardial infarction, and osteoarthritis.

To address KQs 4 and 5, on critiques of reviews and best practices for dealing with clinical heterogeneity, we examined peer and public review comments from three CERs and conducted a literature scan to identify articles that discussed clinical heterogeneity. In addition, we conducted a small number of interviews with key informants to address KQ 5 further and to inform our recommendations for KQ 6.

Results

Determining the answers to two questions—whether an intervention will benefit some patients more (or less) than others and which patients are at greatest (or least) risk of harm when receiving an intervention—is the primary purpose for evaluating clinical heterogeneity in SRs and CERs.

Clinical heterogeneity definitions across review groups. Our first finding was that use and definitions of the terms "heterogeneity," "clinical heterogeneity," and "clinical diversity" varied among review groups, often without any clear distinctions among the definitions. As mentioned earlier, *The Cochrane Handbook for Systematic Reviews of Interventions* defines heterogeneity as "any kind of variability among studies in a systematic review," but states that variability in the participants, interventions, and outcomes studied is termed "clinical diversity." AHRQ, CRD, Cochrane Collaboration, DERP, NICE, and EUnetHTA all discuss variability in the population, interventions, and outcomes. These are three of the six factors to be considered in the development of KQs for AHRQ SRs ("PICOTS," i.e., population, intervention, comparator, outcomes, the timing of their measurement, and setting).

Clinical heterogeneity is closely linked to statistical heterogeneity. The occurrence of clinical heterogeneity may lead to statistical heterogeneity that is detected using techniques such as Cochran's Q test, the I^2 index, or meta-regression. Statistical heterogeneity may *signal* the presence of clinical heterogeneity, methodological heterogeneity, or chance (random error). If reviewers detect statistical heterogeneity, they cannot be sure whether to attribute it to clinical heterogeneity, methodologic heterogeneity, chance, or some combination of the three.

Clinical heterogeneity is also closely related to applicability. This concept, otherwise known as "external validity" or "generalizability," refers to whether, and to what extent, analysts can decide that they can generalize intervention-outcome associations to different persons, treatments, outcomes, or settings.

Addressing how clinical heterogeneity has been handled in the development of key questions. We evaluated the KQs from 123 SRs and CERs conducted by systematic reviewers. We focused on whether the review groups considered demographic variables, disease variables (i.e., disease stage, type, or severity), risk factors for disease, cointerventions, and coexisting conditions.

The groups varied in the extent to which they included demographic variables and disease factors in their KQs. In addition, we detected differences in the extent to which the groups elaborated on how they identified the variables; conditions in which the literature base was more extensive (e.g., hypertension) tended to specify more variables related to clinical heterogeneity. Manuals for these review organizations stressed that reviewers should specify such variables when they develop the protocols for their reviews. Few groups, however, documented exactly when or how they determined which factors to include in their questions.

Report ing on how clinical heterogeneity is handled in the review process. We focused on the results sections from 11 AHRQ reviews. We looked at whether the authors considered demographic and clinical variables during the analysis phase of their work. Of these 11 reviews, all included a clinical factor (disease variable, risk factor, coexisting condition, or cointervention) in their analysis, and 10 considered one or more demographic variables. However, it is important to realize that what EPCs evaluate in their analysis reflects the extent of the available literature and AHRQ does not require specific analyses for investigating clinical heterogeneity.

Five general critiques of how SRs handle clinical heterogeneity were noted. They were from the peer and public review comments for three AHRQ CERs:

1. Missing information on clinically relevant subgroups;
2. Failure to include all studies with relevant information on clinically heterogeneous populations;
3. Too much focus on randomized controlled trials (RCTs) that do not inform "real world" practice;
4. Inappropriate pooling of dissimilar populations or loss of information because of pooling; and
5. Too little discussion of availability and/or evaluation of subgroups in conclusions.

External reviewer comments claimed that the publications failed to address important subgroups, even though all three reviews in fact had had KQs focused on evaluating subgroups. It is very likely that the literature synthesized for these reviews did not have sufficient information to provide summary data on important subgroups or the authors chose not to present subgroup information.

From our literature search, we noted two major concerns. First, timing of subgroup identification is critical (i.e., a priori during the protocol development phase vs. post hoc during the analysis phase). Subgroups identified after the fact are often considered a product of data dredging; these subgroups are likely to be misleading and not confirmed in future studies. Second, the literature also cautions that testing of numerous subgroups without controlling for overall type I error probability may lead to misleading results as well.

Addressing clinical heterogeneity in SRs. We gleaned two best practices from the existing literature. First, authors should identify factors that may cause clinical heterogeneity during the protocol development stage. Second, they should keep the list of factors to as few as possible to avoid misleading results.

We noted similar views from authors of six SRs on myocardial infarction or osteoarthritis. They commented that, ideally, authors should consider such factors during the

protocol development process, but they also acknowledged that, sometimes, too little information about a given topic is available to enable any a priori determination. For that reason, and given the varying literature available for any specific topic when initiating a review, systematic reviewers should be considering clinical heterogeneity throughout the entire review process. This means being attentive to such heterogeneity issues not only during protocol development and analysis of the results, but also as part of developing the inclusion/exclusion criteria, creating abstraction forms, and abstracting data from articles.

Some authors combined clinical and methodological heterogeneity under the rubric of clinical heterogeneity. In addition, many of the publications we reviewed indicated that analysis of individual patient-level data in meta-analyses may allow better assessment of clinical heterogeneity, but the time, cost, and difficulty in obtaining these data are often prohibitive barriers to such analyses.

We provide the following suggestions for extending this work by an evidence-based practice work group. We note, as well, that these are not settled matters in the broader world of systematic reviewers. Thus, any elucidation of these types of questions should prove of benefit beyond the AHRQ EPC ambit. For that reason, and to gain the most up-to-date thinking across many groups dealing with these same problems, AHRQ's EPC program may wish to involve leaders in the SR field from outside AHRQ and outside the U.S.

The 11 questions in Table C, offered in a somewhat "chronological order" as authors might move through a review, are our priority recommendations for what an EPC work group might address. Many might have obvious subquestions, but we believe that this set can establish a robust agenda for any work group.

Table C. Topics for specific charge to the work group

1.	Is the definition of clinical heterogeneity clear enough for future work? Are the distinctions between it and statistical heterogeneity clear as well? Should clinical heterogeneity be distinguished from clinical diversity?
2.	Are the basic categories of clinical heterogeneity introduced for this study—demographic characteristics; clinical variables involving disease severity, stage, or type, risk factors, coexisting conditions, and cointerventions—satisfactory? Are they sufficient?
3.	What process might be developed for determining which clinical heterogeneity factors a review should consider? Would this process differ depending on whether the work (for AHRQ) is a standard systematic review or a comparative effectiveness review?
4.	Should restriction be part of the toolkit for addressing clinical heterogeneity?
5.	Should the process (question 3) mandate "only" a priori statements of clinical heterogeneity factors to be taken in account?
6.	How would such a process take account of what sponsors or nominators of topics have suggested in this context? How would it take elements of the clinical problems, health interventions, and other aspects that differ markedly across the range of reviews that AHRQ sponsors?
7.	Would such a process permit post hoc identification of subgroups for further analysis? If so, what conditions might it set for authors to justify such decisions? What role might individual studies rated "poor quality" (and likely excluded from final analyses) play when no other acceptable evidence on important subgroups exists?
8.	Do appropriate statistical tests exist for assessing clinical heterogeneity? If so, how can such information best be provided as guidance to EPC reviewers?
9.	Should a plan for clinical heterogeneity assessment be part of the posted workplan?
10.	Should each EPC systematic review include a description for how they will handle clinical heterogeneity in the methods section of the review? If so, where should a description of the findings from the clinical heterogeneity assessment be placed?
11.	What recommendations might be made for agreed-upon terminology and standard reporting of clinical heterogeneity results?

Discussion

Clinical heterogeneity exists when patient-level factors—most commonly variables related to patient characteristics, disease location and severity, comorbidities, and accompanying treatment—influence or modify the magnitude of the treatment effect. Unlike statistical heterogeneity which can be quantified, clinical heterogeneity is detected and evaluated without using statistical methods. Moreover, if a reviewer detects statistical heterogeneity, he/she cannot be sure whether to attribute it to clinical heterogeneity, methodologic heterogeneity (study design issues), chance, or some combination of the three. Also, we say there is a distinction between clinical and methodological heterogeneity but drawing a firm line between them is often difficult.

Clinicians and patients are interested in which factors have important effects on the intervention-outcome association because this information helps them understand who is likely to benefit the most, who is likely to benefit the least, and who has the greatest (or preferably least) risk of experiencing adverse outcomes. Because every patient is different with respect to their comorbidities and concurrent treatments, clinicians need to know the extent to which a test or treatment might benefit the next patient they see. Likewise, patients seek to know whether a test or treatment will benefit them individually.

Unfortunately, research studies are not designed to answer treatment questions about individual patients. In order to provide robust conclusions, we need to study large numbers of individuals. And, if we want to be able to say anything about specific subgroups of the population such as those who are over age 65 with very severe disease, we need to make sure that we include enough people who are elderly with late-stage disease in our studies (i.e., we have to power our studies to estimate the intervention-outcome association in this subgroup). Thus, researchers need to consider subgroups in the planning of their original research studies.

This assumes, however, that we know which particular subgroup is important to evaluate in the primary studies that are eventually included in SRs and CERs. For some intervention-outcome associations, the research literature is so sparse that specifying subgroups a priori for an SR or CER is almost impossible because this information has not been published. Those conducting original research studies may have evaluated important subgroups but not included the information in the published paper. This may occur for two reasons: because the study was not powered to provide robust estimates for that subgroup or because including information on all of the subgroups evaluated might be considered data dredging.

Alternatively, systematic reviewers may be faced with the dilemma of having too many subgroups to evaluate if the topic has been well researched. Reviewers need to be cognizant that when evaluating numerous subgroups, either in original research papers or SRs, one might want to control for multiple comparisons or else the findings may be misleading.

Addressing clinical heterogeneity in various types of SRs is a necessary step and some review organizations do provide guidance and rules on how to identify and evaluate clinical heterogeneity. The AHRQ *EPC Methods Guide* does not yet provide guidance on how to identify clinical heterogeneity variables that might modify estimates of treatment outcome in any given review, although an upcoming guidance paper on assessing applicability does provide some suggestions on how to select factors that may be considered for assessing both applicability and clinical heterogeneity. Neither does it discuss methods for addressing clinical heterogeneity or provide suggestions for inclusion in final reports. We conclude that our findings and the recommendations noted earlier can provide a foundation for an AHRQ workgroup to strategize on how to best address these issues for the EPCs.

Evidence Report

Chapter 1. Introduction

Background

The RTI International–University of North Carolina at Chapel Hill (RTI-UNC) Evidence-based Practice Center (EPC) was asked to explore how systematic review groups have dealt with clinical heterogeneity and to determine the best practices for addressing clinical heterogeneity in systematic reviews (SRs) and comparative effectiveness reviews (CERs). Such best practices, to the extent they exist, may enable AHRQ's EPCs to address critiques from patients, clinicians, policymakers, and other proponents of health care about the extent to which "average" estimates of the benefits and harms of health care interventions apply to individual patients or to small groups of patients sharing similar characteristics.

Such users of reviews often assert that EPC reviews typically focus on broad populations and, as a result, often lack information relevant to patient subgroups of particular concern to them. More importantly, even when EPCs evaluate literature on homogeneous groups, there may be varying individual treatment for no apparent reason, indicating that average treatment effect does not point to the *best* treatment for any given individual. Thus, the health care community is looking for better ways to develop information that may foster better medical care at a "personal" or "individual" level. (We do not use the phrases "personalized medicine" or "individualized medicine" here, however, because of the terms' commonly understood applications in genetics and genomics.)

Before focusing on clinical heterogeneity per se, we needed to examine *heterogeneity* in the context of SRs and CERs. This term refers broadly to among-study differences in the effect measure of choice (e.g., an odds ratio, a risk ratio, or a risk difference). Observing heterogeneity in SRs and CERs is important because its appearance suggests differences in, for example, the patients included in the studies with regard to demographics, coexisting conditions, or treatments (or some combination of these elements). Clinicians need to understand which factors are associated with heterogeneity in study results if they are going to be able to apply this information to their individual patients as they strive to practice patient-centered care.[1] We provide a glossary of terms we will be using in this document and our definitions for each term (Table 1).

Table 1. Glossary of terms as used in the report

Applicability	As related to evidence-based practice, is similar to generalizability or external validity of the evidence in an SR or CER; it concerns whether information can be said to pertain directly to a broad selection of patient populations, outcomes, settings, and so forth.
Clinical heterogeneity	Variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect (e.g., odds ratio, risk ratio, risk difference).
Effect measure, measure of effect	A value that measures the effect of a factor on the frequency or risk of a health outcome. Which measure depends on the study design but can be an odds ratio, risk ratio, risk difference, or absolute difference.
Effect-measure modification	Effect-measure modification is said to occur when an intervention-disease association differs according to the level of a factor under investigation. Factors that may influence the intervention-disease association include demographics (age, sex, race/ethnicity), severity of disease, comorbidities, and cointerventions.
Heterogeneity	In the context of SRs and CERs, refers to among-study differences in the effect measure of choice.
Methodologic heterogeneity	In the context of SRs and CERs, refers to among-study differences in the effect measure of choice due to variability and quality of the study design and analysis.
Outcome	A change in health status due to an intervention.
Outcome measure	How the outcome is evaluated (e.g., a validated instrument or clinical assessment for detecting treatment response).
Statistical heterogeneity	Variability in the observed treatment effects beyond what would be expected by random error.

The literature has highlighted heterogeneity as an important construct since the origins of evidence-based medicine in the early 1990s. For years, clinicians have criticized the external validity (generalizability) of randomized controlled trials (RCTs), and the summary of RCTs via SRs and meta-analyses, for not providing treatment information on the patients they typically see in their practices.[2] RCTs, especially those conducted for approval by regulatory agencies, often exclude patients with coexisting conditions and may include only those who have shown a willingness to adhere to treatment using procedures such as "run-in periods."[3] Clinicians know, however, that patients respond differently to drugs within the same therapeutic class and that factors such as age, sex, race, coexisting conditions, and cointerventions may play a role in differential response to treatment.

Comprehensive assessment of how systematic reviewers approach clinical heterogeneity requires a definition of clinical heterogeneity and a discussion of the various types of heterogeneity often present. One major issue in dealing with heterogeneity, and clinical heterogeneity in particular, is that the term has not been used consistently; another is that no definition exists that all researchers endorse or use. Some authors refer to clinical heterogeneity as clinical diversity and others as heterogeneity of treatment effects. For this report, we define "clinical heterogeneity" as the variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect (e.g., odds ratio, risk ratio, risk difference).

Researchers often consider any heterogeneity problematic because it indicates that pooling across studies may not be appropriate, yet true heterogeneity can be informative by

suggesting new avenues for research investigations.[4,5] Ideally, one could differentiate between heterogeneity of treatment effects stemming from factors such as demographics, coexisting conditions, treatments, or genetics (what we and other researchers term "clinical heterogeneity") and that resulting from variability in study design and analysis (which we and others refer to as "methodologic heterogeneity"). However, clinical and methodologic heterogeneity are intertwined; both can and do co-occur in SRs and CERs.

Alternatively, "statistical heterogeneity" refers to the variability in observed treatment effects that is beyond what would be expected by random error (chance), and it is assessed by testing the null hypothesis that the studies have a common treatment effect given a chosen *P*-value. Statistical heterogeneity may *signal* the presence of clinical heterogeneity, methodological heterogeneity, or chance. If reviewers detect statistical heterogeneity, they cannot be sure whether to attribute it to clinical heterogeneity, methodologic heterogeneity, chance, or some combination of the three. Therefore, in setting out to assess clinical heterogeneity, reviewers must be aware that chance and methodologic heterogeneity may distort their findings or conclusions.

The difference between clinical heterogeneity and statistical heterogeneity can be thought of as a cause and an effect relationship, respectively: when clinical heterogeneity is apparent across studies included in a meta-analysis, it can lead to some degree of statistical heterogeneity.[6] For the purposes of this report, what we consider as clinical heterogeneity refers to what can be detected through inclusion and examination of various population subgroups, recognizing that observed differences may reflect a mix of clinical heterogeneity, methodologic heterogeneity, and chance.

Handling clinical heterogeneity in SRs, CERs, and meta-analyses has been an ongoing challenge for the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Program in developing summary estimates in either meta-analyses or qualitative or narrative assessments. Yet, this information is critical for helping to determine which patients will benefit most from an intervention, who will benefit least, and just as importantly, who is at greatest risk of harms from the intervention. Like AHRQ, other organizations, globally, face the same types of challenges. AHRQ's sister program, the DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Network, has recently brought the question of clinical heterogeneity into sharper focus as an important element of comparative effectiveness research.

The issue clinicians face when trying to use SRs, and CERs in particular, for treating their patients, is that the studies included in the reviews may be so diverse in terms of populations, interventions, comparators (when relevant), outcomes, timing, and/or settings (the "PICOTS" paradigm) as to preclude any pooling of data (statistical or otherwise). When this is the case, reviewers cannot draw solid conclusions about the effectiveness of interventions for important subgroups of the population. On the other hand, clinical trials, observational studies, and SRs may be so restricted in eligible or included study populations as to be only marginally relevant for broad application of SR or CER findings or policymaking.

In the first SR from the RTI-UNC EPC, *Pharmacotherapies for Alcohol Dependence,*[7,8] the EPC chose not to conduct a meta-analysis because the interventions and outcomes of the RCTs included in the review were too clinically diverse. For instance, not all studies reviewed required that patients be detoxified before being randomized; in addition, the availability and type of counseling combined with pharmacotherapy differed among studies. Illustrating the narrowness of populations studied, a recently published Drug Effectiveness Review Project (DERP) report on second-generation antidepressants, also conducted by staff of the RTI-UNC

EPC, determined that few studies addressed two important populations: the elderly and those with chronic conditions.[9]

How to address clinical heterogeneity when conducting SRs, CERs, and meta-analyses has been discussed in the literature, but little consensus has been reached on best practices for identifying and understanding the factors underlying such heterogeneity. Some researchers conduct *subgroup analyses* to isolate the factor(s) implicated in such heterogeneity, or they may apply *meta-regression techniques* to summary data or, when available, to individual patient data from the studies being meta-analyzed.[10] Other investigators rely on *restriction*, limiting, perhaps drastically, the range of patient or subject characteristics examined or the measures or study settings included. A case in point is reviews that focus on one particular subgroup (e.g., only persons ages 45 to 54 years); this is essentially a lack of clinical heterogeneity. In this report, we consider reviews to address clinical heterogeneity if they either explore comparisons of therapies for all subgroups for a variable (e.g., age groups such as 45 to 54, 55 to 64, and ≥ 65) or if they restrict the population to more narrowly defined subgroups (e.g., 45 to 54 years only).

AHRQ commissioned the RTI-UNC EPC to conduct this methods project to better understand and identify how AHRQ and the EPC program can better address clinical heterogeneity and concerns of stakeholders related to clinical heterogeneity. Of particular interest is how well clinicians, policymakers, and others can draw conclusions about the effectiveness of interventions from reports that account for clinical heterogeneity in both the populations of interest to them and the populations studied. The findings from this report should be relevant to systematic reviewers in the U.S. and elsewhere.

Key Questions

To address our charge for this methods project, the EPC set out to answer six key questions (Table 2). AHRQ assigned these KQs to us and we worked with AHRQ staff and the EPC program's Scientific Resource Center (SRC) at the Oregon Health & Science University on approaches to address the five empirical issues and their subquestions. KQ 6 asked the project team to put forward ideas that an AHRQ cross-EPC work group might take on in 2010 or later, drawing on our findings for the first five questions. As implied by KQ 6, AHRQ wanted to understand how its EPC program (and the EPC tasks in the Effective Health Care Program related to production of CERs) can better address concerns of stakeholders related to clinical heterogeneity—that is, how confidently clinicians, policymakers, and others can draw conclusions about the effectiveness of interventions from reports that account for clinical heterogeneity in both the populations of interest to them and the populations studied. Although the first set of audiences are oriented to the United States, we believe that our findings can be helpful to systematic reviewers globally. These questions called for us to do the following:

- define clinical heterogeneity;
- determine how SRs and CERs, supported by AHRQ and other groups or agencies internationally, have attempted to address this concept through just the central issues (i.e., KQs) of a given report, with particular attention to the idea of population subgroups;
- determine how SRs or CERs actually have included or examined clinical heterogeneity as part of their overall analyses;
- look into whether public or private "external peer review" of SRs or CERs have brought up issues about clinical heterogeneity or whether any more general discussions of this problem have appeared in the literature;

- determine whether any best practices have been proposed or put into play in SRs or CERs; and, finally,
- put forward ideas that an AHRQ cross-EPC work group might tackle in the next year or so (i.e., 2010 or beyond).

Table 2. Key questions for methods report on clinical heterogeneity

1. What is clinical heterogeneity?
a. How has it been defined by various groups?
b. How is it distinct from statistical heterogeneity?
c. How does it fit with other issues that have been addressed by the AHRQ Methods Manual for CERs?
2. How have systematic reviews dealt with clinical heterogeneity in the key questions?
a. What questions have been asked?
b. How have they preidentified population subgroups with common clinical characteristics that modify their intervention-outcome association?
c. What are best practices in key questions and how these subgroups have been identified?
3. How have systematic reviews dealt with clinical heterogeneity in the review process?
a. What do guidance documents of various systematic review groups recommend?
b. How have EPCs handled clinical hetereogeneity in their reviews?
c. What are best practices in searching for and interpreting results for particular subgroups with common clinical characteristics that may modify their intervention-outcome association?
4. What are critiques in how systematic reviews handle clinical heterogeneity?
a. What are critiques from specific reviews (peer and public) on how EPCs handled clinical heterogeneity?
b. What general critiques (in the literature) have been made against how systematic reviews handle clinical heterogeneity?
5. What evidence is there to support how to best address clinical heterogeneity in a systematic review?
6. What questions should an EPC work group on clinical heterogeneity address?

Some of these (especially KQs 1-5) are "empirical" issues, about which we gathered information from existing SRs and CERs and from interviews. They are covered in Chapter 3 of this report. The remaining KQ (i.e., KQ 6) is derivative of the earlier ones and is more appropriately discussed in Chapter 4. For transparency, we emphasize that the evidence provided in this methods report cannot be considered comprehensive. In particular, because of the extensiveness of the issue and the breadth of organizations around the globe that support development of such reviews (broadly defined), we had to limit our review to systematic samples of the literature.

Organization of This Report

As described more thoroughly in Chapter 2 on methods, we abstracted information from SRs and CERs (some of which may have used meta-analytic techniques) that had been conducted by international organizations involved in literature syntheses (chiefly AHRQ, the Cochrane Collaboration, the Drug Effectiveness Review Project [DERP], the National Institute for Health and Clinical Excellence [NICE], and a variety of others as appropriate to the KQ at hand). We also conducted a literature scan for methods publications addressing clinical heterogeneity, and we conducted discussions with several authors who lead SR efforts for certain reviewer groups.

Chapter 3 provides our results of this data collection, with numerous summary tables, for KQs 1–5. Chapter 4 discusses the results and all KQs, including KQ 6, in more detail. References can be found following Chapter 4. Exact search strings can be found in Appendix A. Appendix B provides materials for the key informant interviews. Appendix C contains three evidence tables and Appendix D has acknowledgments and lists our peer reviewers. Evidence Table C1 concerns KQ 1; Evidence Table C2 (for four reviewer groups and reviews from two

Centre for Reviews and Dissemination databases) concerns KQs 2 and 3; and Evidence Table C3 describes information for KQs 4 and 5.

Chapter 2. Methods

To produce this methods report on issues relating to clinical heterogeneity and the six key questions (KQs) listed in Chapter 1 (Table 2), we used a variety of data sources. For KQ 1, we reviewed guidance documents developed by organizations involved in developing systematic reviews (SRs) and clinical practice guidelines. The literature base for KQs 2 and 3 included selected, relevant literature—that is, SRs, comparative effectiveness reviews (CERs), and similar reports (e.g., health technology assessments). We conducted literature scans to identify articles that discussed clinical heterogeneity and the related concept, effect-measure modification, in SRs and meta-analyses to address KQs 4 and 5. Lastly, we conducted a small number of interviews with key informants to address KQs 5 and 6, which asks for inputs into guidance for a possible work group for the Evidence-based Practice Center (EPC) program of the Agency for Healthcare Research and Quality (AHRQ). The remainder of this chapter documents our methods in detail.

Some experts in evidence-based practice believe that restriction is a way to address clinical heterogeneity while others regard restriction as a way to avoid clinical heterogeneity. Although not apparent from the KQs listed in Table 2, the project team considered restriction as a technique for addressing clinical heterogeneity. As noted in Chapter 1, we use the term "restriction" to refer to any substantially limited subset of a population (e.g., women who are 45 to 64 years of age who have localized breast cancer). Focusing on such a specific subgroup provides information useful for treating women in this category but when a publication with this restricted population is included in an SR or CER, it cannot provide actionable clinical information on patients falling outside the specific subgroup. In this example, a review would not provide direct information on women who also have localized breast cancer but who are either younger than 45 or older than 64.

KQ 1. Definition of Clinical Heterogeneity

We sought and abstracted information from guidance reports and manuals prepared by US and international organizations engaged in preparing SRs and CERs. The target organizations included:

- Agency for Healthcare Research and Quality (AHRQ, http://www.ahrq.gov);
- Centre for Reviews and Dissemination (CRD, http://www.crd.york.ac.uk/crdweb/);
- Cochrane Collaboration (http://www.cochrane.org/);
- Drug Effectiveness Review Program (DERP) of the Oregon Health & Science University (OHSU) Center for Evidence-based Policy (http://www.ohsu.edu/ohsuedu/research/policycenter/DERP/);
- a variety of health technology assessment organizations; this includes those involved with the International Network of Agencies for Health Technology Assessment (INAHTA, http://www.inahta.org/), such as the Agency for Health Technology Assessment in Poland (AHTAPol), the Agencias y Unidades de Evaluaćion de Tecnologías Sanitarias de Madrid (AUETS), the Finnish Office for Health Technology Assessment (Finohta), the Institute for Quality and Efficiency in Health Care (IQWIG), the Australian Safety and Efficacy Register of New International Procedures-Surgical (ASERNIP-S), the Canadian Agency for Drugs and Technologies in Health (CADTH), The Netherlands Organisation for Health Research and Development (ZonMw), and the Agence d'evaluation des technologies et modes d'intervention en sante (AETMIS); and

- National Institute for Health and Clinical Excellence (NICE, http://www.nice.org.uk/).

Working with AHRQ agency personnel, RTI project staff identified guidance documents and manuals prepared by national and international organizations as "instructions" for their staff or participants in the conduct of SRs and health technology assessments. We then reviewed and abstracted information from these documents to address KQ 1.

We captured information on whether the various manuals addressed clinical heterogeneity at all and, if so, how they defined the term. To address KQ 1b about its distinction from statistical heterogeneity, we evaluated whether manuals defined statistical heterogeneity and whether they explored or provided examples about the relationship between clinical and statistical heterogeneity, including the use of restriction as a way to handle clinical heterogeneity. Based on this information, we developed a summary table of definitions (Table 1) and discuss these in more detail in Chapter 3. Evidence Table C1 (Appendix C) documents the information abstracted from these materials.

KQs 2 and 3. Clinical Heterogeneity in Key Questions and Systematic Reviews

Selection of Publications To Review

KQ 2s and 3 are focused on how those conducting SRs and CERs (or any constituent meta-analyses) handle clinical heterogeneity in developing their KQs (KQ 2) and how they deal with clinical heterogeneity during the abstraction and synthesis (i.e., analysis) process (KQ 3). We derived evidence for these two KQs from a systematic sample of SRs, CERs, and meta-analyses that had been completed by four organizations with extensive expertise in literature syntheses: AHRQ, Cochrane Collaboration, DERP, and NICE. Our sample also included syntheses that were catalogued in the Centre for Reviews and Dissemination's Database of Abstracts of Reviews of Effects (DARE) and the Health Technology Assessment (HTA) database. NICE documents may actually be full clinical practice guidelines, the basis for which will be an SR and possibly meta-analyses within it; the other research groups are unlikely to produce practice guidelines but rather confine their reports to SRs or CERs (and constituent meta-analyses in some cases).

Given the number of SRs, CERs, and meta-analyses published to date, we could not review all of them to assess how their KQs and analyses might have addressed clinical heterogeneity. In planning our sampling strategy, we faced the question of whether to conduct a more in-depth analysis that compared only a few disease areas or a less comprehensive analysis of many disease areas. Focusing on too few topic areas was thought to be insufficient for generalization and not likely to provide the varying perspectives that AHRQ was seeking. By contrast, focusing too broadly was simply beyond the resources and time available for the project.

In consultation with the AHRQ Task Order Officer and the Director of the Scientific Resource Center (SRC), the project team developed a sampling strategy that allowed comparisons of methods used by the SR groups while "controlling for" irrelevant differences by making these comparisons within the same disease areas. In our judgment, for example, comparing methods used by one SR group on osteoarthritis with approaches applied by a second SR group report on myocardial infarction would be inappropriate; the factors that would need to

be evaluated (e.g., age groups, sex, disease severity, comorbidities, and outcome measures) would not necessarily be the same across these two conditions.

For this reason, we developed a sampling process that allowed us to make two types of comparisons: (1) the content and treatment of clinical heterogeneity in SRs and similar reports *across* review groups for a given clinical condition and (2) those topics across clinical conditions *within* a review organization. The project team decided to seek a condition for each major body system and three different types of cancer (breast, lung, prostate). Table 3 lists the final set of 15 conditions on which we focused. By including reviews from AHRQ, Cochrane, DERP, and NICE and then systematically sampling syntheses from CRD's DARE and HTA databases, we included reviews conducted, globally, by many different review teams.

Table 3. Conditions selected for detailed review

• Breast cancer
• Lung cancer
• Prostate cancer
• Cesarean section (limited to cesarean section on maternal request)
• Chronic kidney disease
• Chronic obstructive lung disease
• Depression
• Dyspepsia (generally dealt with in reviews of gastroesophageal reflux disease)
• Heart failure (including but not limited to congestive heart failure)
• Heavy menstrual bleeding
• Hypertension
• Labor induction
• Myocardial infarction
• Irritable bowel syndrome
• Osteoarthritis

Databases and Search Strategies for Condition-Specific Reports

An important caveat for this report is that our searches for KQs 1-5, although systematic, cannot be considered comprehensive as we are describing the current state of the literature on how clinical heterogeneity is handled in SRs, CERs, and meta-analyses. The specific approaches are listed below; unless otherwise noted, we sought completed reviews for each of the 15 topics in Table 3. Table 4 lists the numbers of reviews we found from Cochrane and CRD sources.

Table 4. Number of unduplicated reports identified from the Cochrane Library, DARE, and HTA databases

	Database Source		
Disease or Condition	**Cochrane**	**CRD**	**Combined**
Breast cancer	85	138	65
Lung cancer	84	56	22
Prostate cancer	23	42	18
Cesarean section	16	6	8
Chronic kidney disease	69	27	9
Chronic obstructive pulmonary disease	59	59	30
Depression	314	200	130
Dyspepsia/GERD	37	19	9
Heart failure (or congestive heart failure)	43	28	9
Heavy menstrual bleeding	13	5	5
Hypertension	218	174	79
Irritable bowel syndrome	12	17	13
Labor induction	27	13	6
Myocardial infarction	115	176	89
Osteoarthritis	41	64	28

Cochrane=Cochrane Collaboration; CRD=Centre for Reviews and Dissemination Databases which includes the DARE database, Database of Abstracts of Reviews of Effects and HTA database, Health Technology Assessment Database); GERD=gastroesophageal reflux disease

For the AHRQ reviews, we searched the Evidence-based Practice Center section of AHRQ's website (http://www.ahrq.gov/clinic/epcix.htm).

To identify Cochrane reviews for these 15 conditions, we searched the Cochrane Library (http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0) on June 23, 2009, using the broad "condition" term (e.g., breast cancer), limiting to the years 2007-2009.

For determination of which DERP reports to include in this review, we searched the DERP website (http://derp.ohsu.edu/about/final-products.cfm). Because DERP reports focus on medications to treat specific indications, we were able to find relevant reports only for eight of the 15 conditions (congestive heart failure, chronic obstructive pulmonary disease [COPD], depression, gastroesophageal reflux disease [GERD], hypertension, irritable bowel syndrome [IBS], myocardial infarction, and osteoarthritis.

For the NICE reports, we conducted a search similar to that for the AHRQ reports. On the NICE website (http://www.nice.org.uk/Guidance/Topic), we looked for the topic of interest and then selected the most recent review. For example, for hypertension, two reviews were available: *Hypertension: Management of Hypertension in Adults in Primary Care* published in June 2006 and *Management of Type 2 Diabetes - Management of Blood Pressure and Blood Lipids* published in October 2002; we selected the former. Similarly, for myocardial infarction, we could select "The clinical effectiveness and cost effectiveness of drugs for early thrombolysis in the treatment of acute myocardial infarction" published in October 2002 or "Post myocardial infarction: secondary prevention in primary and secondary care for patients following a myocardial infarction" published in May 2007; we chose the latter.

For the health technology assessments and reviews listed in the DARE database, we searched the CRD databases. The HTA and DARE databases contain details of completed and ongoing SRs, meta-analyses, and health technology assessments produced by review organizations; we focused on completed assessments. We used a search strategy similar to that for the Cochrane reviews, but using the terms appropriate for searching the CRD database:

(CONDITION) NOT ("provisional abstract" OR "cost study" OR "in process"), limiting to those completed between 2007 and 2009. We conducted these searches on June 22, 2009, and netted the number of citations shown in Table 4. Reviews that were identified from the DARE and HTA databases are not specific to any one review group. They provide a general view of how clinical heterogeneity has been dealt with, with the sole caveat that the HTA reviews address only health technology assessments.

In some cases, the same reviews may appear in both the Cochrane Library database and the CRD database, so we merged all 30 searches into one EndNote library. After the consolidation, we reviewed the titles and abstracts to identify duplicates. Using the search function in EndNote, we subset the reviews into topic-specific "custom groups" based on whether the condition appeared in either the title or the abstract of the citation. Table 4 presents these combined, unduplicated, and condition-focused counts of reports (N = 1,114 from these two sources).

Eventually, we had to establish a sampling strategy so that we would have an approximately equal number of reviews for each topic area (disease or condition) for KQs 2 and 3. In addition, we wanted the distribution of the reviews that we were including to reflect, proportionally, the number of reviews done for that specific topic by each review group. These constraints resulted in our having to identify between 5 and 12 reviews per disease area.

To identify the specific completed studies to review, we designed a sampling strategy (for the EndNote library) that focused on SRs, CERs, or meta-analyses from only the Cochrane or CRD database (DARE and HTA only) searches (i.e., those in Table 4). From the EndNote database that held the topic-specific citations from the Cochrane and CRD (DARE and HTA) searches, we used a random number generator to randomly select approximately 10 percent of the reviews for conditions in which we had identified 90 or more reviews, 30 percent for topics for which we identified 30 or so reviews, 50 percent for topics for which we identified between 6 and 8 reviews, and 100 percent when we had 5 or fewer reviews. For example, for heart failure (including congestive heart failure), we identified nine reviews (seven from the DARE search, one from Cochrane, and one that was an HTA review; see right-hand column in Table 4). Because we had only one review on this condition from both Cochrane and the HTA database, we selected both. We then needed to randomly select three from among the seven identified from the DARE search, and in this case we focused specifically on congestive heart failure (not simply heart failure). For diseases such as breast cancer or depression, we had many more reviews overall and by the different groups, and the sampling strategy was then more complex (or random).

We provide the actual number of AHRQ, Cochrane, DERP, and NICE reviews as well as the number of reviews selected from the DARE and HTA database searches in Table 5. Note that these reviews reflect a broad swath of how clinical heterogeneity has been addressed globally; collectively this still cannot be construed as a comprehensive review of the entirety of the SR literature.

Data Abstraction Process

We abstracted information on how these reviews handled clinical heterogeneity when developing their own core questions and when analyzing their findings. Some sources use the phrase "key questions," and others do not, but the basic concept is the same.

Detailed information can be found in Evidence Table C2 in Appendix C, for each of the six sources of reviews separately. The Evidence Table lists the "key questions" for each review

and whether the authors addressed clinical heterogeneity or subgroups as part of their core questions (i.e., the information to address KQ 2). For reviews that did address clinical heterogeneity in their main questions (i.e., the information to address KQ 2), we abstracted information on the factors the authors evaluated, categorized into the following groups: demographics (e.g., age, sex, race, ethnicity), disease severity (including stage for cancers) and site (e.g., hip or knee for osteoarthritis), risk factors (e.g., smoking), cointerventions, coexisting conditions, and consideration of pregnancy.

Similarly, we abstracted information on whether the authors addressed clinical heterogeneity during the analysis phase of their work. Again, we recorded information, insofar as available, on how they addressed clinical heterogeneity and which factors were evaluated for each KQ. When available, we provide more specific information on the factors evaluated during the analysis phase rather than the six broad categories we used for our KQ 2 evaluation.

Table 5. Number of reports reviewed to address key questions 2 and 3

	Number of Reports Selected for Review by Research Group						
Disease or Condition	**AHRQ**	**Cochrane**	**DARE**	**DERP**	**HTA**	**NICE**	**Total**
Breast cancer	1	2	2	NA	3	1	9
Lung cancer	1	3	3	NA	1	0	8
Prostate cancer	1	2	1	NA	2	0	6
Congestive heart failure	1	1	3	3	NA	0	8
Cesarean section	1	2	2	NA	NA	0	5
Chronic kidney disease	NA	2	2	NA	NA	0	4
Chronic obstructive pulmonary disease	1	4	5	2	NA	0	12
Depression	1	5	4	1	4	1	16
Dyspepsia/GERD	1	NA	3	1	1	1	7
Heavy menstrual bleeding	NA	4	NA	NA	NA	0	4
Hypertension	1	3	2	1[a]	3	0	10
Irritable bowel syndrome	NA	3	4	1	NA	0	8
Labor induction	1	2	NA	NA	NA	0	3
Myocardial infarction	NA	3	3	2[a]	3	1	12
Osteoarthritis	1	3	3	1	2	1	11
All conditions	11	39	37	12[b]	19	5	123

AHRQ=Agency for Healthcare Research and Quality; Cochrane=Cochrane Collaboration; DARE=Database of Abstracts of Reviews of Effects; DERP=Drug Effectiveness Review Program; GERD=gastroesophageal reflux disease; HTA=health technology assessment organizations; NA=not applicable because no report was available from the research group; NICE=National Institute for Health and Clinical Excellence.

[a]3 additional reports were reviewed but were duplicates of those relating to heart failure

[b]18 total reports (12 unique) reviewed; 6 reports were duplicates among heart failure, hypertension, and myocardial infarction

KQs 4 and 5. Critiques and "Best Practices"

To understand external comments or critiques of how SRs or CERs handled the concept of clinical heterogeneity (KQ 4) in their analyses and, then, to identify best practices for handling clinical heterogeneity within SRs and CERs (or constituent meta-analyses) (KQ 5), we carried out several different tasks. Specifically, we conducted a formal literature search on these issues, carried out a citation search on three important articles, and analyzed comments from external peer reviewers and public comments on three specific AHRQ-supported SRs or CERs.

Literature Search and Citation Analysis

Literature search. In our first step, we conducted a MEDLINE® literature search using PubMed that focused on methods related to conducting reviews and meta-analyses. The intent of this search was to identify whether guidance on the conduct of SRs and CERs (1) differentiated among different types of heterogeneity and (2) described how to identify factors causing clinical heterogeneity, including evaluating particular subgroups and conducting analyses on individual patient data rather than using the summary results from publications. We conducted our search on May 1, 2009, using the search terms provided in Appendix A to address these issues with a net result of 1,065 articles.

Citation analysis. Because we were concerned that our literature search might not target the publications of greatest value for the project, we also used the Science Citation Index (http://thomsonreuters.com/products_services/science/science_products/a-z/science_citation_index) to identify publications that had cited three seminal articles about the importance of evaluating clinical heterogeneity in meta-analyses. These were:

- Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *American Journal of Epidemiology* 1995;142:383-387.[5]
- Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *American Journal of Epidemiology* 1995;142:371-382.[11]
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal.* 1994;309:1351-1355.[6]

In all, 389 publications cited one or more of these three publications. We added these citations to our EndNote database. This yielded a final total of 1,432 references for addressing KQs 4 and 5.

Other sources. We also worked with the lead librarian for the SRC to identify publications from their methods EndNote library that provided information on clinical heterogeneity or subgroup analysis.

Included and excluded articles. For all the entries in our compiled EndNote library with publications from the MEDLINE search, the Science Citation Index search, and suggestions from the SRC library, we conducted a title and abstract review. The aim was to determine which citations focused specifically on issues related to clinical heterogeneity rather than simply reporting on an SR or meta-analysis without commenting on clinical heterogeneity or subgroups. We conducted a further, full-text review of those articles that appeared to have useful information on handling clinical heterogeneity and retained those that in fact did have such information.

We then abstracted information from these included publications in Evidence Table C3 (Appendix C). Although this report concerns clinical heterogeneity as its focus, our research revealed that heterogeneity in general has been classified into different categories (clinical; methodological; and statistical). We will define these categories in Chapter 3. We described these factors in a binary (yes/no) fashion.

Peer Reviewer and Public Reviewer Comments

For KQ 4a, we worked with the SRC to make available the peer and public reviewer comments from three CERs: *Comparative Effectiveness of Drug Therapy for Rheumatoid Arthritis and Psoriatic Arthritis in Adults*;[12] *Comparative Effectiveness of Percutaneous Coronary Interventions and Coronary Artery Bypass Grafting for Coronary Artery Disease,*[13] and *Comparative Effectiveness of Second-generation Antidepressants in the Pharmacologic Treatment of Adult Depression*,[9] We did not select these three topics randomly. We knew the depression review had numerous peer and public comments. We selected the cardiovascular and arthritis reviews because these were the conditions selected for our key informant interviews.

We identified specific comments that addressed clinical heterogeneity. In particular, we determined whether such peer or public reviews

- Noted that the CER had (or had not) discussed the information available on a specific subgroup (e.g., for rheumatoid arthritis);
- Questioned whether the report had focused on individuals with a specific type of disorder (e.g., coronary artery disease);

- Identified instances in which the authors had done any inappropriate analysis such as grouping mild and severe disease subgroups (e.g., coronary artery disease); or
- Commented on the paucity of data for important subgroups such as those with physical or neurovegetative comorbidities (e.g., depression).

The information derived from evaluating the peer and public review comments of these three CERs is provided in a summary table in Chapter 3.

Key Informant Interviews

The project team anticipated that it would find little guidance on how authors of SRs, CERs, and meta-analyses determined the factors to evaluate in their reviews that might shed light on subgroups of importance, disease severity, or comorbidities. To address this expected gap, we conducted six key informant interviews in September and October 2009 with the authors of SRs on osteoarthritis and myocardial infarction that had been developed by different review groups.[14-19] The purpose of the interviews was to explore how researchers specified aspects of clinical heterogeneity or how they dealt with the related concept of effect-measure modification in their main, core questions or how they handled this issue in their analyses. We were limited to six informants for time and funding constraints. We also used these findings to inform KQs 2–5.

Participant Selection

Participants were eligible to participate based on authorship of an SR pertaining to osteoarthritis or myocardial infarction accessed through literature databases and abstracted for inclusion in Evidence Table C2. Before conducting the interviews, RTI submitted an application along with the interview guide to RTI's Institutional Review Board and received exemption status.

Using lead author email addresses provided in the relevant osteoarthritis and myocardial infarction reviews, the RTI project director sent an introductory email to potential participants explaining the purpose of the study and interview task, and requesting their participation. Upon receiving confirmation of participation, an RTI staff member emailed the participant to set up a date and time for the interview. If we did not receive any response to the initial email within 3 days, the project director sent a follow-up email. If no response was received after the second email, RTI staff conducted an Internet search for a telephone number and/or alternative email address for the lead author. If the lead authors were unavailable to participate, RTI staff attempted to contact the secondary authors. One (sometimes two) RTI project team members facilitated the interviews with each author by telephone.

Procedure and Analysis

The interviews lasted approximately 30-45 minutes and addressed the specific review the authors were involved with as well as questions about their handling of clinical heterogeneity more generally. More specifically, the questions covered authors' approaches to study protocol development, opinions on formulation of subgroups for an SR, handling of clinical heterogeneity in the KQs and analysis, use of manuals and guidance during study protocol development, and additional considerations in the selection of patient or disease factors (Appendix B has the interview questions). Interviews were recorded, transcribed, and compared with written notes taken during the interviews. Staff used QSR NVivo 8 (http://www.qsrinternational.com/), a

qualitative software package, to auto-code the interview data by question and to facilitate a question-by-question analysis. Because of the lateness of the interviews relative to the project's conclusion, interview notes could not be reviewed for clarification by interviewees after transcription.

Chapter 3. Results

As noted in Chapter 1, we address here the five (of six) key questions (KQs) for which we had some empirical information relating to systematic reviews (SRs), comparative effectiveness reviews (CERs), or meta-analyses. Table 2 listed the full set of KQs. We provide summary tables of primary findings here; the three evidence tables pertaining to these KQs can be found in Appendix C. Chapter 2 described our various reviews of the literature for different KQs and documented the yields from those searches. It also explains how we conducted the key informant interviews.

KQ 1. What Is Clinical Heterogeneity?

The focus of this report is on best practices for addressing clinical heterogeneity within SRs. Ideally, if SRs were able to provide summary effect estimates that would differentiate between patients who would benefit from an intervention in contrast to those who would either not benefit or who might be harmed, then this would allow clinicians to provide treatment tailored to their patients. Thus, clinical heterogeneity should be valued because it helps inform patient care. *The Cochrane Handbook for Systematic Reviews of Interventions*[20] defines heterogeneity as "any kind of variability among studies in a systematic review," but defines clinical heterogeneity as variability in the participants, interventions and outcomes studied.

The term "heterogeneity" as used in the epidemiology literature and assessed in clinical studies refers to an intervention-disease association that differs according to the level of a factor under investigation. The term "effect-measure modification" is often used to clarify that heterogeneity can be observed on the relative scale, the absolute scale, neither, or both, and may be present on one scale but not the other (hence, it is the specific *effect measure* where the heterogeneity is observed).

The presence of effect-measure modification may suggest a biologic (or etiologic) effect of a factor upon the intervention-disease relationship, or it may reflect one or more biases. A factor can modify an effect measure for the intervention-disease relationship when baseline rates of the disease vary among factor subgroups or when the baseline rates do not vary among those subgroups. However, it is important to note that baseline rates may vary within subgroups of a factor whether or effect-measure modification is observed on any scale. This is because whether a given factor modifies baseline risk of disease and whether or not it modifies the effect of a particular treatment on that disease, or the direction or degree to which it modifies that treatment effect, are unrelated. Many different clinical factors can be evaluated as influencing the intervention-disease association (i.e., as modifiers of one or more effect measures), including demographics (age, sex, race/ethnicity), severity of disease, disease risk factors, coexisting diseases, and cointerventions. Many, but not all such factors, influence baseline rates of disease as well.

Ideally, expert advice and the prior literature should be used during the protocol development stage to identify factors that may impact the heterogeneity of treatment effects. Often, however, subgroup analyses in trials either are not defined a priori or are done inadequately, leading to false-positive findings because of multiple statistical tests having been conducted or false-negative findings because of lack of power.[21]

Nevertheless, clinical knowledge is constantly evolving, and the impact of heterogeneity on treatment effects may be unknown at the design stage of a trial. Post hoc subgroup analyses, therefore, have an important role in research but should be viewed as hypothesis generating and

not as an assessment of an associative relationship. The strength of SRs and CERs in regard to heterogeneity is that, because they review multiple studies on the same intervention, they offer a new opportunity to explore reasons for varying study results.[5]

Evaluating whether there is heterogeneity of the treatment effect in an SR or CER is one of the first steps in an analysis because it is linked to the effect being studied. The fact that clinical heterogeneity is present is a finding to be reported because it helps identify who benefits the most, who benefits the least, and who has the greatest risk of experiencing adverse outcomes. These are central concerns for most users of SRs and CERs because clinicians do not treat "average" patients; they want to know the extent to which a test or treatment might benefit the next patient they see. Thus, more information on the treatment effect across diverse groups of patients may assist clinicians' work and improve the quality of care they can render.

KQ 1a. Definitions of Clinical Heterogeneity by Various Groups

To provide an overview of approaches and definitions of various international institutions, we reviewed the methods manuals from nine organizations or public-sector agencies that produce SRs (or clinical practice guidelines in which SRs are embedded). They are located in the United States (three organizations) and abroad (two from the United Kingdom, one each from Germany and Australia, and two European or global enterprises):

- U.S. Agency for Healthcare Research and Quality (AHRQ, http://www.ahrq.gov),[22]
- Oregon Health & Science University Drug Effectiveness Review Project (DERP, http://www.ohsu.edu/ohsuedu/research/policycenter/DERP),[23]
- HuGENet (Human Genome Epidemiology Network, http://www.cdc.gov/genomics/hugenet/default.htm),[24]
- The (United Kingdom) Centre for Reviews and Dissemination at the University of York (CRD, http://www.york.ac.uk/inst/crd/),[25]
- The National Institute for Health and Clinical Excellence (NICE, http://www.nice.org.uk),[26]
- The German Institute for Quality and Efficiency in Health Care (http://www.nhmrc.gov.au/ IQWIG [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen], http://www.iqwig.de/),[27]
- The Australian National Health and Medical Research Council (NHMRC, http://www.nhmrc.gov.au/),[28]
- The Cochrane Collaboration (http://www.cochrane.org/),[20] and
- The European Network for Health Technology Assessment (EUnetHTA, http://www.eunethta.net).[29]

We summarize here the range of definitions and recommendations about clinical heterogeneity found in their methods manuals.

Of the nine methods manuals reviewed, only five—AHRQ's EPC program, CRD, Cochrane Collaboration, DERP, and EUnetHTA—provided explicit definitions of clinical heterogeneity.[20,22,23,25,29] AHRQ, Cochrane, and CRD used the term "clinical diversity" rather than "clinical heterogeneity." Their manuals have defined "clinical diversity" as variability of study population characteristics, interventions, and outcomes, with "differential response to an intervention" as another way to refer to differences in treatment effects on specific outcome measures because the underlying effect does differ by one of these factors. Table 6 lists the main definitions from these five organizations.

Table 6. Definitions of clinical heterogeneity by five organizations

Review Group	Definition
AHRQ	Variability in study population characteristics, interventions, and outcomes
CRD	Differences in participants, interventions, or outcome measures
Cochrane	Variability in the participants, interventions, and outcomes studied (also termed "clinical diversity")
DERP	Variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies
EUnetHTA	Differences in participant characteristics, interventions, or outcome measures

AHRQ=Agency for Healthcare Research and Quality; CRD=Centre for Reviews and Dissemination; Cochrane=Cochrane Collaboration; DERP=Drug Effectiveness Review Project; EUnetHTA=European Network for Health Technology Assessment.

Cochrane Collaboration definition. The Cochrane manual provides the most detailed discussion of what we are referring to as "clinical heterogeneity," which they define as "the variability in the participants, interventions, and outcomes studied." Variability in participants include personal characteristics (e.g., age, sex, ethnicity) and disease severity and progression, varying baseline risks of experiencing certain events, coexisting conditions, past treatments, and other factors. Differing interventions refer to varying dosages, cointerventions, and surgical techniques; the concept also means inconsistent control interventions (e.g., placebo or active controls).

Cochrane clearly distinguishes clinical heterogeneity from methodological heterogeneity by defining methodological heterogeneity as "the variability in study designs and risk of bias."[30] With respect to methodological heterogeneity, the Cochrane manual contends that differences in methodological factors such as adequate randomization, allocation concealment, and use of blinding among studies will lead to differences in observed treatment effects. Such differences, however, do not necessarily indicate that the true intervention effect varies. Empiric studies have shown that poor study design can lead to an overestimation of the magnitude of the effect.[31,32]

In short, true clinical heterogeneity exists when patient-level factors—most commonly variables related to patient characteristics, comorbidities, and accompanying treatment—may influence or modify the magnitude of the treatment effect.

AHRQ and CRD definition. Like Cochrane, CRD and AHRQ acknowledge that some variation in treatment effects among studies always arises by chance alone. However, if clinical heterogeneity influences the estimated intervention effect beyond what is expected by chance alone, then clinical heterogeneity becomes important. The AHRQ *EPC Methods Guide* lists common examples of factors contributing to clinical heterogeneity: age, sex, disease severity, site of lesion, evolving diagnostic criteria, changes in standard care, time-dependent care, differences in baseline risks, and dose-dependent effects.

Other organizations. The DERP and EUnetHTA manuals use a definition similar to AHRQ, CRD, and Cochrane, but DERP also uses the term "qualitative heterogeneity." No other manuals explicitly define clinical heterogeneity. Their chapters about heterogeneity deal primarily with statistical heterogeneity and the consequences of statistical heterogeneity with respect to meta-analyses.

"Restriction" as a related concept. Cochrane and CRD both caution against "restriction" (i.e., constraining enrollment of subjects, study settings, or what measures to use) as a way of addressing clinical heterogeneity because, they argue, doing so limits the applicability (see below) of the information to patient populations with the condition of interest. Any restrictions with respect to specific population characteristics should be based on a sound rationale.

No other manual addresses restriction. However, "applicability" is sometimes considered a further related concept. Applicability, as related to evidence-based practice, can be thought of as generalizability or external validity of the evidence in an SR or CER; it concerns whether information can be said to pertain directly to a broad selection of patient populations, outcomes, settings, and so forth. The AHRQ *EPC Methods Guide*[22] does address questions of applicability as a characteristic of bodies of evidence. A recent publication on grading the strength of evidence also discusses applicability.[33] We provide more information on applicability in SRs in the discussion section (Chapter 4).

KQ 1b. Distinctions Between Clinical and Statistical Heterogeneity

In contrast to how clinical heterogeneity was defined for KQ 1a, statistical heterogeneity refers to the variability in the observed treatment effects that is beyond what would be expected by random error (chance). Assessing statistical heterogeneity involves testing the null hypothesis that the studies have a common treatment effect given a chosen *P*-value. Clinical heterogeneity can result in statistical heterogeneity.[6]

Authors of SRs have to put forward convincing arguments that clinical heterogeneity did or did not occur when evaluating an outcome of interest in a given review. When clinical heterogeneity is detected, the onus is on the author to determine whether this finding is clinically relevant. In other words, systematic reviewers have to determine whether differences in population characteristics among studies can lead to clinical heterogeneity that could change clinical decisions.

For example, in a CER on treatments for rheumatoid arthritis (RA), the relative benefit of biologic treatments over methotrexate was smaller in patients with early RA than in patients who had long-lasting RA and had failed to respond to other disease-modifying antirheumatic drugs (DMARDs).[12] Although study protocols, drug dosages, follow-up periods, and methodological rigor were very similar between the two sets of trials, the differing stages of RA in the study populations may have produced a substantial variation in the magnitude of treatment effects.

Such an assessment is not always so straightforward. Historically, the impact of clinical heterogeneity has been both under- and over-estimated, based on flawed subgroup analyses or anecdotal clinical evidence.[21] Exploring the impact of clinical heterogeneity in SRs, therefore, has to involve both clinical understanding and formal statistical tests.

Investigating the extent of variation of among-study results is an important part of any SR. Results of a careful assessment provide the foundation from which one or another of two clinically important conclusions can be drawn:

1. Treatment effects are similar statistically despite clinical heterogeneity. Such a finding is an important corroboration of the applicability of study results to more diverse clinical populations.
2. Treatment effects exhibit variation beyond what would be expected by chance alone as indicated by a statistical test. Such a result requires careful investigation of the reasons for and the magnitude of the variation of treatment effects. Findings from such an investigation might then dictate choice of the statistical model for meta-analyses, employment of sensitivity analyses to determine the effect of the variation on the overall pooled estimate, subgrouping of studies to estimate separate pooled estimates by subgroup, or a decision to forego any meta-analysis that pools data inferentially across studies.

Formal statistical methods to assess heterogeneity. Inevitably, even with thoughtfully defined eligibility criteria and well-formulated, focused KQs, studies included in SRs will differ in various ways and will exhibit some variation in treatment effects. This is to be expected, by chance alone (random error). The underlying rationale of statistical tests to assess heterogeneity is to investigate whether existing variations in treatment effects go beyond what would be expected by chance fluctuations alone.

Various statistical methods exist to determine and quantify the degree of variation. Commonly used statistical tests are Cochran's Q test,[34] I^2 index,[35] and meta-regression.[10] Table 7 summarizes common statistical approaches to test for heterogeneity.

Table 7. Summary of common statistical approaches to test for heterogeneity

Cochran's Q test Cochrane's Q test is an extension of the McNemar test that provides a method for testing for differences between three or more matched sets of frequencies or proportions. Cochran's Q test is the traditional test for heterogeneity in meta-analyses. Based on a chi-square distribution, it generates a probability that, when large, indicates larger variation across studies rather than within subjects within a study. The underlying null hypothesis assumes that the true treatment effect is the same across studies and variations are simply caused by chance. A limitation of Cochran's Q test is that it might be underpowered when few studies have been included or when event rates are low. Therefore, it is often recommended to adopt a higher *P*-value (rather than 0.05) as a threshold for statistical significance when using Cochran's Q test to determine statistical heterogeneity.[30,36]
I^2 index The I^2 index is a more recent approach to quantify heterogeneity in meta-analyses. I^2 provides an estimate of the percentage of variability in results across studies that is due to real differences and not due to chance. The I^2 index measures the extent of heterogeneity by dividing the result of Cochran's Q test and its degrees of freedom by the *Q*-value itself. When I^2 is 0%, variability can be explained by chance alone. If I^2 is 20%, this would mean that 20% of the observed variation in treatment effects cannot be attributed to chance alone. Some underlying factor may be the potential effect-measure modifier. An I^2 of less than 25% is usually viewed as low heterogeneity, between 25% and 50% as moderate, and over 50% as high heterogeneity. The limitation of I^2 is that it provides only a measure of global heterogeneity but no information for the factor causing heterogeneity, similar to Cochran's Q test. Meta-regression or subgroup analyses can help determine which factors are causing heterogeneity after the I^2 has been conducted.[37]
Meta-regression Meta-regression models strive to control for and explain differences in treatment effects in terms of study covariates. A meta-regression can be either a linear or a logistic regression model, and it can be based on a fixed or random effects regression.[10] Most commonly, the unit of the analysis is the individual study included in a systematic review or meta-analysis. Predictors in the regression model are study-level characteristics such as study-level location, sample size, length of followup, drop-out rates, or study quality characteristics. In exploring heterogeneity, the advantage of meta-regression is that it determines which study-level characteristics account for heterogeneity, rather than just providing an estimate of the global heterogeneity. Therefore, meta-regression is most commonly used to explore existing heterogeneity. An a priori analysis protocol should be used when meta-regression is applied to avoid spurious results. Nevertheless, meta-regression using study-level characteristics can only partially address issues of heterogeneity. Patient-level characteristics should not be used when individual patient data are not available. Meta-regression analyses of mean patient characteristics from trials (e.g., mean age, mean disease severity) can provide misleading results for individual patients, which is known as the ecological fallacy.[38]

Relationship between clinical and statistical heterogeneity. In SRs, clinical and methodological heterogeneity across studies is often present, regardless of whether treatment effect is measured on the relative or absolute scale.[39] Also, it is possible for one effect measure to be homogeneous and another to be heterogeneous. More problematic is that this heterogeneity is not always measured in its full detail because of incomplete descriptions of intervention protocols, populations, and outcomes. Moreover, it can, but does not always, result in detectable statistical heterogeneity (i.e., variation in treatment effect beyond that expected by chance alone).

Thus, an overall test of heterogeneity may be nonsignificant but a specific aspect of the study populations may be significantly associated with study findings.

Clinical and statistical heterogeneity are closely intertwined. Understanding this relationship is important because they do not have a linear relationship. In other words, high clinical heterogeneity does not always cause statistical heterogeneity and it is critical to realize that statistical heterogeneity can be caused by either or both methodological and clinical heterogeneity.

Common reasons for statistical heterogeneity include the following:

1. Methodological heterogeneity. This can refer to variability in study design, study conduct, outcome measures, and study quality (internal validity). It concerns differences in methodological quality that lead to variations in bias. Empiric studies have shown that high risk of bias often leads to an overestimation of the magnitude of the effect. Such methodological issues could include problems with randomization, allocation concealment, drop-out rates, or statistical analyses (e.g., intention-to-treat vs. per-protocol analyses).[31]
2. Chance. Individual studies, particularly studies with small sample sizes or low event rates, can exhibit extreme results based simply on chance. Such outliers can cause statistical heterogeneity.
3. Biases. In addition to biases that threaten the validity of individual studies and that are captured under methodological heterogeneity, various other biases, including funding and reporting (publication) biases, may cause variability in treatment effects estimated across studies.[6,40] For example, small trials with nonsignificant findings have a higher risk of remaining unpublished than small trials showing significant (or very large) effects.

Consequently, for systematic reviewers assessing heterogeneity, the relationship between clinical and statistical heterogeneity is not always straightforward. Table 8 outlines the different relations between clinical and statistical heterogeneity under the assumption that random error, methodological heterogeneity, and biases do not play a role. When both clinical and statistical heterogeneity are present, the reviewers must consider whether the differences in treatment effect may be due to clinical variability or methodological characteristics. Thus, in some cases, reviewers have to pay close attention to the methods of each study.

The "possible underlying situation" (right column) explains what inferences might be drawn and whether reviewers need to examine the situation further. Figures 1, 2, and 3 illustrate different underlying situations graphically.

Exploration of statistical heterogeneity. As outlined in Table 8, statistical heterogeneity can be present with or without clinical heterogeneity and can be caused by reasons other than clinical heterogeneity. SR authors might be tempted to overinterpret apparent relationships between statistical heterogeneity and clinical variations based on results at hand. Particularly when findings are caused by chance, searching for causes can be misleading.[6] The problem is similar to that of subgroup analyses.[21] Therefore, systematic reviewers must carefully and cautiously explore the reasons for statistical heterogeneity and view results as exploratory rather than causal.

Figure 1. Clinical heterogeneity is present but has a minimal impact on the treatment effect

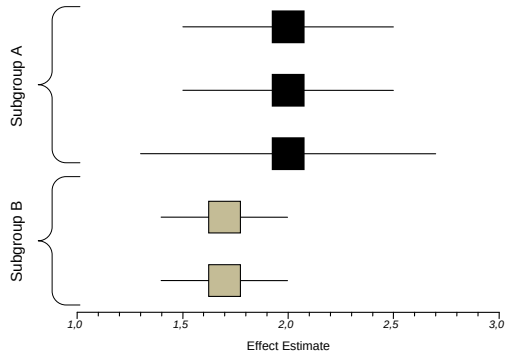


Figure 2. Clinical heterogeneity is present but the relevance of the impact has to be determined on clinical grounds

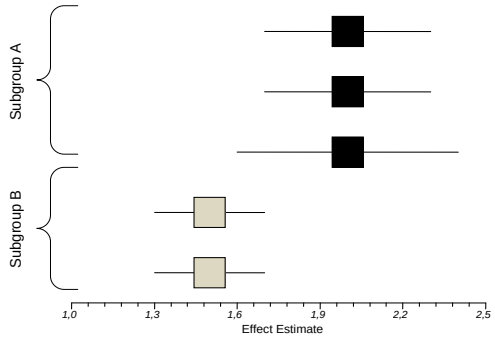
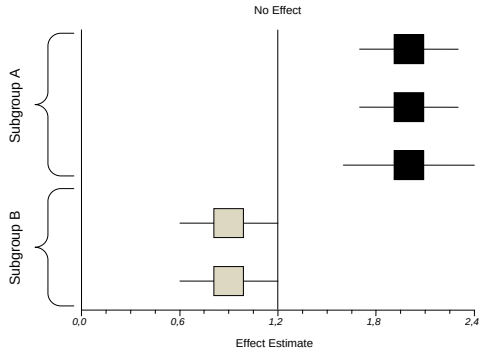


Figure 3. Clinical heterogeneity is present and leads to a clinically relevant impact on the treatment effect (reversed direction)



False conclusions about clinical heterogeneity based on statistical heterogeneity can be summarized in two ways:

1. False-positive conclusion (type I error). The presence of statistical heterogeneity is attributed to clinical differences rather than random variation or confounding.
2. False-negative conclusion (type II error). The presence of statistical heterogeneity is attributed to other factors such as methodological heterogeneity or chance because no clinical heterogeneity is apparent. In reality, unidentified or "not obvious" factors cause variability of the treatment effect. A not-obvious factor might involve items that are important but not measured (or not easily measurable), such as socioeconomic status or genetic makeup. Generally, reviewers use one or more of three common approaches to explore heterogeneity:

Stratified analyses of homogenous subgroups. We distinguish between subgroup analysis and sensitivity analysis using the Cochrane Collaboration Glossary of terms as our basis (http://www.cochrane.org/resources/glossary.htm). Subgroup analysis is an "analysis in which the intervention effect is evaluated in a subset" of particular study participants, or as defined by study characteristics. So for example, the subgroup might be defined by sex (men vs. women), or by study location (urban vs. rural setting). Subgroup analysis tends to be defined a priori, that is, as part of the study protocol.

Table 8. Summary of relationships between clinical and statistical heterogeneity

Clinical Heterogeneity	Statistical Heterogeneity	Possible Underlying Situations
None (populations appear to be similar)	**None**	• Clinical heterogeneity does not truly exist or is not measurable in the available studies. • The evidence is insufficient to draw conclusions as to whether clinical heterogeneity leads to differences in the size of the effect.
None (populations appear to be similar)	**Present**	• Unidentified or unknown clinical heterogeneity is present and needs to be explored. • Methodologic heterogeneity may be causing statistical heterogeneity. • Variations in effects are a consequence of an inappropriate choice of an effect measure.
Present (populations differ in various characteristics that could act as modifiers of the effect measure)	**None**	• Lack of power of statistical tests for heterogeneity produces false-negative result. • Clinical heterogeneity has no impact on the treatment effect. • Clinical heterogeneity has an impact on the treatment effect but the magnitude of the impact is small and of unclear clinical relevance; confidence intervals overlap widely. (See Figure 1)
Present (populations differ in various characteristics that could act as modifiers of the effect measure)	**Present**	• Clinical differences lead to variations in treatment effects; the relevance of the variation has to be determined on clinical grounds. (See Figure 2) • Methodologic heterogeneity may be causing statistical heterogeneity alone or in conjunction with clinical heterogeneity. • Clinical differences modify the effect measure; differences in effects are statistically significant, and both clinically important (e.g., reversed in direction) and relevant. (See Figure 3)

The Cochrane manual advises that subgroup analysis should be tested via an interaction test, not by comparing *P*-values.[20]

Meta-regression. These types of analyses, as discussed earlier, enable investigators to explore sources of heterogeneity in terms of study-level covariates. They must be done with due attention to potential pitfalls and challenges, however.

Sensitivity analyses. Sensitivity analysis is defined as analysis used to "assess how robust the results are to assumptions about the data and the methods that were used." Generally such analyses are post hoc, that is, during the analysis phase of the study. For example, a sensitivity analysis might be conducted to determine if changing study inclusion/exclusion criteria changes the conclusions substantially, or to assess if methods for imputing missing data impact the results. Due to its post hoc nature, sensitivity analysis should be considered exploratory, not confirmatory. Both subgroup and sensitivity analyses are constrained practically and inferentially in terms of the availability of studies and sample size. Both may be subject to the challenge of multiple comparisons.

KQ 1c. Clinical Heterogeneity and Other Issues in the AHRQ Methods Manual

For systematic reviewers, especially those doing CERs in the context of guidance from AHRQ through the Evidence-based Practice Center (EPC) program's *Methods Guide*,[23] identifying potential effect-modifying clinical characteristics is important from the planning stages of the review to the synthesis of the evidence. Specifically, systematic reviewers should consider which factors may be associated with effect-measure modification at all stages of the review: from framing the KQs, through protocol development when inclusion and exclusion criteria are determined, in the development of the abstraction forms, analysis, and finally, when the data are summarized either qualitatively or quantitatively. However, assessing heterogeneity is not an explicit part of the workplan template that EPCs are presently expected to follow.

As outlined above, clinical heterogeneity can be the cause of statistical heterogeneity. Systematic reviewers who consider combining studies statistically must explore existing statistical heterogeneity. If clinical heterogeneity is suspected to be the cause of statistical heterogeneity, researchers might abstain from meta-analyses because populations across different trials might be too different to be combined in a meaningful meta-analysis. Even if statistical heterogeneity does not appear to be present, suspicion of clinical heterogeneity may be cause to limit meta-analysis. The distinction will be important to clinicians. If clinical heterogeneity is confirmed, it may change clinical decisionmaking with individual patients.

As mentioned earlier, clinical heterogeneity is also closely related to a broader issue of SRs and CERs: namely, the assessment of the applicability of findings and conclusions. "Applicability" has been defined as inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments, and outcomes.[41] For many audiences in the broader world of health services research, policy research, or quality improvement and patient safety evaluations, this concept is often equated with generalizability or external validity.

Deciding to whom findings of SRs apply requires a close understanding of which patient groups benefit the most and which the least from a given medical intervention. Any specific intervention is unlikely to benefit everyone equally, even with a statistically significant and clinically relevant overall treatment effect. A hypothetical intervention with a number needed to treat (NNT) of 3 to achieve a beneficial outcome would be considered highly effective.[42] Nevertheless, in this scenario, two of three treated patients would not experience any benefit from the intervention. Moreover, they might even experience harm from the treatment with no gain or benefit. To identify those who benefit the most or the least is an important piece of information when available; with this information clinicians can appropriately tailor treatments to individuals.

In turn, being aware of treatments for which clinical heterogeneity is not a significant issue is also important. A common criticism of SRs is that they provide average results that are not applicable to individual patients with varying risks and prognostic factors. To identify treatments that are not or only minimally affected by clinical heterogeneity can lead to a more rational use of interventions and help avoid both over- and under-treatment.

Input from experts and stakeholders is important to identify issues of clinical heterogeneity and to frame applicability issues.[22] These experts can provide insights into typical health care practice. Numerous studies have reported important differences between patients enrolled in trials and those treated with the same condition in everyday practice.[43-45]

Whether such differences translate into varying treatment effects remains generally unclear in many areas. Some treatment- and condition-specific knowledge, however, can be gained from the exploration of clinical heterogeneity in SRs and CERs.

KQ 2. How Have Systematic Reviews Dealt with Clinical Heterogeneity in the Key Questions?

KQs 2a and b. Key Questions and Pre-Identified Subgroups

KQs 2a and 2b addressed how the various research groups dealt with clinical heterogeneity in their KQs (KQ 2a) and how they identified (a priori) population subgroups of interest (KQ 2b). We note below the distribution of reviews with respect to including demographic variables and addressing disease variables (i.e., disease stage, type, severity, or site) or similar clinical variables in KQs, as well as pre-identifying population subgroups based on clinical characteristics. Results are presented by each of the four research groups included in this study and the reviews identified from CRD's DARE and HTA abstracts database. Of interest were the following 15 clinical conditions: breast cancer, lung cancer, prostate cancer, cesarean section, chronic kidney disease, chronic obstructive pulmonary disease (COPD), depression, dyspepsia, heart failure (including congestive heart failure), heavy menstrual bleeding, hypertension, irritable bowel syndrome (IBS), labor induction, myocardial infarction, and osteoarthritis. The reviews completed for each research group and selected from the DARE and HTA databases were listed in Table 4 (Chapter 2).

Agency for Healthcare Research and Quality (AHRQ). To address KQs 2a and 2b for AHRQ, we obtained their SRs for 11 medical conditions: breast cancer, lung cancer, prostate cancer, heart failure, cesarean section, COPD, depression, dyspepsia, hypertension, labor induction, and osteoarthritis (Table 9).[46-56] No AHRQ SRs were available for chronic kidney disease, heavy menstrual bleeding, IBS, or myocardial infarction.

Cochrane Collaboration. To address KQs 2a and 2b for the Cochrane Collaboration, we obtained SRs from Cochrane for 14 medical conditions (39 reviews in all): breast cancer, lung cancer, prostate cancer, heart failure, cesarean delivery, chronic kidney disease, COPD, depression, heavy menstrual bleeding, hypertension, IBS, labor induction, myocardial infarction, and osteoarthritis (Table 10).[14,16,18,57-92] No Cochrane SRs were available for dyspepsia.

Database of Abstracts of Reviews of Effects (DARE). To address KQs 2a and 2b for SRs located in CRD's DARE database, we identified and obtained SRs for 12 medical conditions (37 reviews in all): breast cancer, lung cancer, prostate cancer, cesarean delivery, chronic kidney disease, COPD, depression, heart failure, hypertension, IBS, myocardial infarction, and

osteoarthritis (Table 11).[15,17,93-127] No SRs for dyspepsia, heavy menstrual bleeding, or labor induction were available from the DARE database.

Table 9. AHRQ's use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Population Subgroups Pre-Identified (Yes/No)? If so, how?
Breast cancer[46]	Age	RF	No
Lung cancer[47]	No	DV, CI, CC	No
Prostate cancer[48]	Age, race/ethnicity	CC, DV	No
Heart Failure[49]	Age, race, sex, income level	CC	No
Cesarean section[50]	Race/ethnicity, sex (fetal), socioeconomics	RF, CC	No
COPD[51]	No	No	NA
Depression[52]	Age, race/ethnicity, sex	DV, CI, CC	No
Dyspepsia[53]	Unspecified	RF, CI	No
Hypertension[54]	Age, race/ethnicity, Sex	CI, CC	No
Labor induction[55]	No	RF	No
Osteoarthritis[56]	Age, race, sex	CI, CC	No

No indicates that the report did not consider the variable for the key questions.
NA indicates that the variable was not applicable because KQs did not address subgroups.

Table 10. Cochrane Collaboration use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Were population subgroups pre-identified? If so, how?
Breast cancer[57,58]	No	No	NA
Lung cancer[59,61]	No	No	NA
Cadona Zorilla et al., 2008[60]	No	DV	No
Prostate cancer[62,63]	No	No	NA
Cesarean delivery[65,66]	No	No	NA
Chronic kidney disease[67]	No	CI	No
Roderick et al., 2007[68]	No	No	NA
COPD[69-71]	No	No	NA
Yang et al., 2007[72]	No	DV	No
Depression[73-77]	No	No	NA
Heart failure[64]	Age	No	No
Heavy Menstrual Bleeding[78-81]	No	No	NA
Hypertension[85,86]	No	No	NA
Wiysonge et al., 2007[82]	Age, race/ethnicity	No	No
Hodson et al., 2007[83]	Age	DV	No
Abalos et al., 2007[84]	Sex	DV, CC	No
Evans et al., 2007[87]	Age	No	No
Labor induction[88]	No	DV	No
Boulvain et al., 2008[89]	No	No	NA
Myocardial infarction[16,18,90]	No	No	NA
Osteoarthritis[14,91,92]	No	DV	No

No indicates that the report did not consider the variable for the key questions.
NA indicates that the variable was not applicable because key questions did not address subgroups.

Table 11. DARE use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Were population subgroups pre-identified? If so, how?
Breast cancer[93,94]	No	No	NA
Lung cancer[95]	No	DV, RF	No
Micames et al., 2007[96]	No	DV	No
Coory et al., 2008[97]	No	No	NA
Prostate cancer[98]	No	No	NA
Heart Failure[99-101]	No	No	NA
Cesarean section[102]	No	RF	No
Press et al., 2007[103]	No	NA	NA
Chronic Kidney Disease[104]	No	No	NA
Strippoli et al., 2008[105]	No	DV	No
COPD[106,108,110]	No	DV	No
Niesink et al., 2007;[107] Singh et al., 2008[109]	No	No	NA
Depression[111,113]	Age	No	No
Cuijpers et al., 2008;[112] Barbui et al., 2008[114]	No	No	NA
Dyspepsia[115,116]	No	No	No
Wang et al., 2007[117]	No	DV	No
Hypertension[118]	No	No	NA
Connell et al., 2008[119]	Race	No	Based on US and UK statistics
IBS[120,121,123]	No	No	NA
Ford et al., 2009[122]	No	RF	No
Myocardial infarction[99-101]	No	No	NA
Ioannidis and Katritsis, 2007[125]	No	DV	No
Osteoarthritis[15]	No	No	NA
Minns et al., 2007[126] Pisters et al., 2007[127]	No	DV	No

No indicates that the report did not consider the variable for the key questions.
NA indicates that the variable was not applicable because KQs did not address subgroups.

Drug Effectiveness Review Project (DERP). To address KQs 2a and 2b for DERP, we obtained SRs for eight medical conditions: COPD, depression, dyspepsia, heart failure, hypertension, IBS, myocardial infarction, and osteoarthritis. We randomly selected 18 of their SRs; however, six reports were duplicates (among heart failure, hypertension, and myocardial infarction) (Table 12).[128-139] No DERP SRs were available for breast cancer, lung cancer, prostate cancer, cesarean section, chronic kidney disease, heavy menstrual bleeding, or labor induction.

Table 12. DERP use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Were population subgroups pre-identified? If so, how?
COPD[131,132]	Age, race, sex	RF, CC, CI, DV	No
Depression[133]	Age, race, sex	CI, CC	No
Dyspepsia[134]	Unspecified	CI, CC	No
Heart Failure[128-130]	Age, Race, Sex	CI, CC	No
Hypertension			
Chou et al., 2005[128]	*Duplicate under Heart Failure		
Furmaga et al., 2006[129]	*Duplicate under Heart Failure		
Helfand et al., 2009[130]	*Duplicate under Heart Failure		
McDonagh et al., 2005[135]	Age, race, sex	CI, CC	No
IBS[135,136]	Age, race, sex	CI, CC	No
Myocardial Infarction			
Dailey et al., 2007[137]	Age, race, sex	CI, CC	No
Helfand et al., 2006[138]	Age, sex, other; unspecified in another key question	CC, CI, RF	No
Helfand et al., 2009[130]	*Duplicate under Heart Failure		
Chou et al., 2005[128]	*Duplicate under Heart Failure		
Furmaga et al., 2006[129]	*Duplicate under Heart Failure		
Osteoarthritis			
Chou et al., 2006[139]	Unspecified	CI, CC	No

Health Technology Assessment Database from CRD. To address KQs 2a and 2b from the Health Technology Assessment (HTA) database, we obtained SRs for eight medical conditions: breast cancer, lung cancer, prostate cancer, depression, dyspepsia, hypertension, myocardial infarction, and osteoarthritis (Table 13).[140-149] There were no SRs for cesarean section, chronic kidney disease, COPD, heart failure, heavy menstrual bleeding, IBS, or labor induction completed during the 2007-2009 time period.

National Institute for Health and Clinical Excellence (NICE). To address this question for NICE, we reviewed five SRs—one each on breast cancer, depression, dyspepsia, myocardial infarction, and osteoarthritis. We were able to identify NICE reviews for all of the conditions, but due to time and resource constraints, we were only able to focus on five SRs (Table 14).[150-154]

KQ 2c. "Best Practices" for Key Questions

We used the manuals reviewed for KQ 1 to address KQ 2c. In contrast with our discussion above for KQs 2a and 2b, which provides our findings by review group, we do not carry through with this format below.

Only the DERP manual recommends explicitly that investigators develop a KQ on patient subgroups. The context is the need to assess whether the comparative effectiveness or tolerability and safety of drugs vary in patient subgroups defined by demographics (age, racial groups, sex or gender, or similar factors), use of other medications, or presence of coexisting conditions. This advice is not couched in "scoping" terms.

Table 13. HTA use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Were population subgroups pre-identified? If so, how?
Breast cancer			
Adelaide Health Technology Assessment, 2008[140]	Age	No	No
Korencan et al., 2007[141]	No	No	NA
Dunfield and Severn, 2007[142]	No	RF	Based on the Gail[155] and BRCAPRO models[156]
Other cancers			
Adelaide Health Technology Assessment, 2007[143]	No	No	NA
Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen, 2007[144]	No	DV	No
Pearson et al., 2007[145]	Unspecified	No	No
Depression[146-149]	No	No	NA
Dyspepsia			
Swedish Council on Technology Assessment in Health Care, 2007[157]	No	No	NA
Hypertension[158-160]	No	No	NA
Myocardial Infarction[161,162]	No	No	NA
National Institute for Health and Clinical Excellence, 2007[163]	*Duplicated under NICE		
Osteoarthritis			
Canadian Agency for Drugs and Technologies in Health, 2007[164]	No	No	NA
Samson et al., 2007[19]	Age, Race/ethnicity, Sex	RF, DV	No

No indicates that the report did not consider the variable for the key questions.
NA indicates that the variable was not applicable because KQs did not address subgroups.

Table 14. NICE use of clinical heterogeneity in key questions

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC), Cointerventions (CI)	Were population subgroups pre-identified? If so, how?
Breast cancer[154]	No	DV, CI,	No
Depression[150]	Age, sex	DV ,RF, CC,	Review group consensus
Dyspepsia[151]	Unspecified	RF, CC	No
Myocardial Infarction[152]	Age, ethnicity	DV, CC, CI	Literature-based; other existing guidance
Osteoarthritis[153]	Unspecified	Unspecified	Review group consensus

Three groups appear to suggest one or another method of "scoping" KQs. AHRQ recommends one approach—namely, performing a preliminary search for relevant trials and the consultation of experts in the field. In this context AHRQ recommends that authors focus carefully on all aspects of the review questions to ensure that they specifically examine subgroups of interest in their review. CRD suggests considering factors that may be investigated for subgroup analysis, including participants' age, sex or gender, socioeconomic status, ethnicity, and geographical area; disease severity; and presence of any comorbidities, before any KQs are stated. Finally, NICE recommends convening a scoping workshop before KQs are formulated to identify which patient or population subgroups should be specified (if any).

The Cochrane handbook does not explicitly discuss subgroups during the process of formulating the KQs; it deals with subgroup analysis in the data analysis chapter. Nevertheless, the Cochrane handbook does discuss restriction with respect to specific population characteristics or settings during the formulation of KQs; this might be regarded as a way to lay out the scope of the issues insofar as clinical heterogeneity is concerned. It specifically advises that authors should consider any relevant demographic factors and notes (as mentioned for KQ 1, above) that any restriction should be based on a sound rationale because restriction limits the applicability of SRs.

No other manual provides guidance on how to address clinical heterogeneity in KQs.

KQ 3. How Have Systematic Reviews Dealt With Clinical Heterogeneity in the Review Process?

For this KQ, we summarized recommendations from the guidance documents we abstracted for KQ 1and provided best practices from these documents. Although this report focuses on addressing clinical heterogeneity in the KQs (KQ 2) and in the analysis phase (KQ 3), we did not find guidance documents, studies, or commentaries indicating that clinical heterogeneity *must be considered* at all stages of the review, from its inception with forming the KQs, developing the inclusion and exclusion criteria, designing the abstraction form, abstracting the information, and then analyzing the findings, and synthesizing the results.

Besides reviewing the guidance documents from KQ 1, we also identified whether AHRQ EPCs considered clinical heterogeneity during the analysis phase of their reviews.

KQ 3a. Recommendations from Guidance Documents

Agency for Healthcare Research and Quality. The AHRQ *EPC Methods Guide* recommends that biological or clinical factors that may influence the occurrence of clinical heterogeneity in the treatment effect be determined a priori based on previous reviews or expert opinion. Then, when framing the KQs for the review, the authors can develop the questions to include the factors contributing to clinical heterogeneity or suggest subgroup analyses to explore these factors in the analysis. With respect to handling clinical heterogeneity in analyses, the manual advises that when it is present, the authors should explain the issues that they considered, including the range of differences in clinical factors that would be considered acceptable for pooling, in deciding whether or not to combine studies using meta-analysis. Any meta-analyses should include sensitivity analyses. The *Methods Guide* does not address restriction as a way of addressing clinical heterogeneity. Currently, assessing heterogeneity is not part of the workplan template.

Cochrane Collaboration and Centre for Reviews and Dissemination. The Cochrane Collaboration and CRD are the only institutions that provide guidance on how to assess clinical heterogeneity and how to deal with clinical heterogeneity in SRs. Both manuals recommend assessing the importance of clinical heterogeneity by visually exploring differences in the magnitudes of treatment effects as a first step. This approach requires plotting point estimates with confidence intervals on a common scale for each study. A forest plot, as used for meta-analysis, would probably be the most appropriate graph. Both institutions recommend investigating the overlap of confidence intervals. If confidence intervals do not overlap or overlap only to a small degree, more formal statistical methods (e.g., chi-square tests) should be considered.

Specifically, the Cochrane handbook suggests that authors consider subgroup analyses as well as meta-regression for addressing clinical heterogeneity. However, meta-analysis will be informative and appropriate only if the study participants, interventions, and outcomes are sufficiently homogeneous. It also provides guidance on the use of restriction with respect to specific population characteristics or settings.

As Cochrane reviews are intended to be widely relevant internationally, the manual advises that authors must justify exclusion of studies based on population characteristics using a sound rationale and must explain this in their review. For example, focusing a review of the effectiveness of prostate cancer screening on men between 50 and 60 years of age may be justified on the basis of biological plausibility, previously published SRs, and existing controversy. By contrast, authors should avoid focusing a review on a particular subgroup based on age, sex, or ethnicity when no underlying biologic or sociological justification can be found for doing so, as this would increase the likelihood of type 1 error. When reviewers are uncertain whether effects among various subgroups of people may differ in important ways, they may be best advised to include all the relevant subgroups and then test for important and plausible differences in the analysis (see Chapter 9, Section 9.6 of the handbook). Subgroup analyses should be planned a priori, stated as a secondary objective, and not driven by the availability of data.

The CRD manual suggests that investigators explore clinical heterogeneity using subgroup analyses that are planned during protocol development. However, when authors cannot plan for subgroups a priori because little information is available at the protocol development stage, they should use an adaptive process with the process specified in the protocol. (The developers of the manual do not provide an example of what an adaptive process might look like.) When authors plan to use restriction, CRD advises that the restrictions put in place should be clinically justifiable such that the results are relevant to the population of concern.

Drug Effectiveness Review Program. DERP guidance does not distinguish among clinical, methodologic, and statistical heterogeneity; rather it discusses heterogeneity in general. Authors of DERP reviews are instructed to consider heterogeneity using the populations, interventions, comparators, outcomes (PICO) framework to determine whether meta-analysis is appropriate. The guidance states that reviewers should use qualitative summaries when meta-analysis is not appropriate. The DERP guidance does not discuss use of restriction for addressing clinical heterogeneity.

Other organizations. The EUnetHTA guidance provides no guidance on how to address clinical heterogeneity; it does indicate that authors note whether clinical heterogeneity is present.[29] Whether clinical heterogeneity is present can be conveyed using tables specifying the populations, interventions, settings, and outcome measures. EUnetHTA does not include restriction as a way of dealing with clinical heterogeneity.

The HuGENet handbook addresses clinical heterogeneity through use of subgroups based on disease or sociodemographic characteristics. Authors should clearly specify subgroups. Details of the subgroup analysis can be provided in text rather than in tabular format unless the subgroup analysis was pre-specified as a primary issue to be evaluated in the review. HuGENet does not include restriction as a way of dealing with clinical heterogeneity.

IQWIG focuses on subgroups as a way to evaluate consistency of treatment results across populations and subgroups such as gender and baseline disease risk. They do not discuss restriction as a means of handling clinical heterogeneity.

Finally, neither the NHMRC guidance nor the NICE manual makes recommendations about exactly how to handle clinical heterogeneity in analyses, and neither discusses restriction.

KQ 3b. Evidence-based Practice Center Practices for Clinical Heterogeneity

KQ 3b asked how AHRQ's EPCs have dealt with the concept of clinical heterogeneity in their SRs and CERs. To address this question, we sought SRs (including CERs) from AHRQ for all 15 medical conditions noted earlier. Because AHRQ requested a broad review of conditions, the principal investigator for this study selected one condition to represent each body system.

Of these, we obtained reviews on 11 conditions: breast cancer, lung cancer, prostate cancer, cesarean delivery, COPD, depression, dyspepsia, heart failure, hypertension, labor induction, and osteoarthritis. In addition, as no AHRQ report dealt with dyspepsia, we used an SR for gastroesophageal reflux disease (GERD) instead. We selected one SR or CER for each of the 11 conditions, counting dyspepsia, regardless of how many reviews EPCs might have been completed for a given topic over the years.[46-56] No AHRQ SR was available for chronic kidney disease, heavy menstrual bleeding, IBS, or myocardial infarction.

We note the distribution of reviews with respect to whether they included demographic variables and addressed disease variables (i.e., disease stage, type, severity, or site) or similar clinical variables (Table 15).

Table 15. Use of demographic or disease variables in AHRQ systematic reviews

Condition and Citation	Demographic Variables	Disease Variables (DV), Risk Factors (RF), Coexisting Conditions (CC). Cointerventions (CI)
Breast cancer [46]	Age	DV, RF
Lung Cancer [47]	NA	DV
Prostate cancer[48]	Age, race/ethnicity	DV, CC
Heart failure[49]	Age, race/ethnicity, sex	CC
Cesarean delivery[50]	Age, race/ethnicity	RF, CI
COPD[51]	Race/ethnicity, sex,	DV. RF
Depression[52]	Age, race/ethnicity, sex	DV, RF, CC, CI
Dyspepsia (GERD)[53]	Age, sex	DV
Hypertension[54]	Age, race/ethnicity, sex	CC, CI
Labor induction[55]	Age,	DV, RF
Osteoarthritis[56]	Age, race/ethnicity, sex	CC, CI

KQ 3c. "Best Practices" for Considering Intervention-Outcome Associations

This subquestion pertained to all organizations considered for KQ 1, not just AHRQ. We comment in detail below only if a manual or handbook provided some explicit advice about analyses or statistical tests to be used in examining associations between interventions and treatment outcomes taking clinical heterogeneity into account.

Cochrane Collaboration and Centre for Reviews and Dissemination. The Cochrane manual suggests that authors determine, at the point of writing their protocols, which characteristics may

be associated with clinical heterogeneity so they can develop a plan to assess these factors during the analysis; the manual also suggests consideration of meta-regression. An initial step when studies reflect inconsistencies is to evaluate whether statistical heterogeneity exists. However, because the power of the heterogeneity test for detecting clinical heterogeneity is low, they suggest using the I^2 with a *P*-value of 0.10 as this *P*-value can provide the strength of the available evidence. When evaluating forest plots, authors should consider the overlap in confidence intervals.

The CRD guidance suggests that authors should examine forest plots, chi-square tests (*Q*-statistic), and the I^2 test as a means of assessing whether clinical heterogeneity is influencing the treatment effect.

Drug Effectiveness Review Program. The DERP manual suggests that reviewers consider whether there are differences in the patient populations, interventions, and outcomes and if the studies are of similar quality before determining whether a meta-analysis should be performed. When meta-analyses are inappropriate, the data should be summarized qualitatively.

Human Genome Epidemiology Network. Clinical heterogeneity with respect to intervention-outcome associations is not addressed specifically by the HuGENet handbook. However, it does advise that heterogeneity in general can be assessed in one or more ways: the estimate of among-study variance (I^2 statistic) and meta-regression with sensitivity analyses.

German Institute for Quality and Efficiency in Health Care. The IQWIG manual did not address clinical heterogeneity specifically although it does provide guidance on assessing heterogeneity in general. Their guidance suggests a priori determination of possible effect-measure modifiers that might affect the treatment-outcome association in particular patient subgroups. Studies that are strongly heterogeneous may be meta-analyzed only when the reasons for the heterogeneity are plausible and justifiable. The extent of heterogeneity should be quantified using the I^2 statistic.

Other organizations. AHRQ does not make any explicit recommendation regarding how authors should assess whether clinical heterogeneity affects the intervention-outcome relationships in its SRs or CERs but does provide guidance on the possible choices for its evaluation. The EUnetHTA guidance has no recommendation for assessing whether clinical heterogeneity influences the intervention effect. The NHMRC manual discusses heterogeneity only in general but suggests that authors should explore possible causes of variation in outcome estimates even when the test for heterogeneity is not statistically significant. Finally, the NICE manual also does not specifically address clinical heterogeneity with respect to outcome estimates or effects. It states, however, that authors should describe and justify their meta-analytical techniques and approaches. This guidance includes specifications for any subgroup analyses and sensitivity analyses.

KQ 4. What Are Critiques in How Systematic Reviews Handle Clinical Heterogeneity?

KQ 4a. Critiques from Peer and Public Reviews of AHRQ Evidence-based Practice Center Reports

As with KQ 3a, this issue related only to CERs from AHRQ EPCs. It specifically deals with external peer review and public comments for three *draft* CERs from AHRQ EPCs:

- Comparative Effectiveness of Drug Therapy for Rheumatoid Arthritis and Psoriatic Arthritis in Adults[12]
- Comparative Effectiveness of Percutaneous Coronary Bypass Grafting for Coronary Artery Diseases[13]
- Comparative Effectiveness of Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression.[52]

All had had KQs addressing subgroups, the most relevant of which are reproduced in Table 16.

Table 16. Clinical heterogeneity variables specified in key questions for AHRQ comparative effectiveness reviews

CER Topic	Key Question and Specific Clinical Heterogeneity Variables
Comparative Effectiveness of Drug Therapy for Rheumatoid Arthritis and Psoriatic Arthritis in Adults	KQ 4. What are the comparative benefits and harms of drug therapies for rheumatoid arthritis and psoriatic arthritis in subgroups of patients based on stage of disease, history of prior therapy, demographics, concomitant therapies, or comorbidities?
Comparative Effectiveness of Percutaneous Coronary Bypass Grafting for Coronary Artery Diseases	KQ 2. Is there evidence that the comparative effectiveness of PCI and CABG varies based on: • Age, sex, race, or other demographic risk factors? • Coronary disease risk factors, diabetes, or other comorbid disease? • Angiographic-specific factors including, but not limited to, the number of diseased vessels amenable to bypass or stenting, vessel territory of stenoses (e.g., left main or anterior descending coronary arteries, right coronary artery, circumflex coronary artery), diffuse vs. focal stenoses, left ventricular function, or prior revascularization procedures? • CABG-specific factors including, but not limited to, cardiopulmonary bypass mode (normothermic vs. hypothermic), type of cardioplegia used (blood vs. crystalloid), or use of saphenous vein grafts, single or bilateral internal mammary artery grafts, or other types of bypass grafts? • Clinical presentation (e.g., stable angina or unstable angina based on New York Heart Association functional class I-IV, acute coronary syndrome, cardiogenic shock, acute myocardial infarction with or without ST elevation, or silent ischemia)? • Adjunctive medical therapies, such as short-term intravenous or oral antiplatelet drugs, or long-term use of oral antiplatelet drugs? • Process characteristics such as provider volume, hospital volume, and setting (e.g., academic vs. community)? • Prior PCI or CABG revascularization procedures?
Comparative Effectiveness of Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression	KQ 3. Do medications or combinations of medications (including tricyclics in combination) used to treat depression differ in their efficacy or effectiveness for treating accompanying symptoms, such as anxiety, insomnia, and neurovegetative symptoms? • Do medications differ in their efficacy and effectiveness in treating the depressive episode? • Do medications differ in their efficacy and effectiveness in treating the accompanying symptoms? KQ 5. How do the efficacy, effectiveness, or harms of treatment with antidepressants for a depressive syndrome differ for the following subpopulations? • Elderly or very elderly patients; • Other demographic groups (defined by age, ethnic or racial groups, and sex); • Patients with medical comorbidities (e.g., ischemic heart disease, cancer); • Patients with psychiatric and behavioral comorbidities (e.g., substance abuse disorders); and • Patients taking other medications

CABG=coronary artery bypass grafting; PCI=percutaneous coronary intervention.

The source of these comments was compilations provided by the SRC. We did not review the Peer Review Disposition Reports (PRDRs) that all EPCs produce to indicate how they dealt with peer or public comments in their final reports. Thus, comments noted below may have been accurate (leading to revisions to the final CER) or inaccurate or irrelevant (meaning that the final reports did not have any related revisions); in all cases, however, the PRDR would have had an explanation of the disposition made for all comments.

Table 17 presents our summary synthesis by type of comments or concerns that either independent peer reviewers or public commentators made about these draft reports. All three reports were criticized for lacking either information on the clinically relevant subgroups or clarity on which comparisons were being made. One reviewer suggested that "to avoid confusion in the interpretation of [the] analysis, it must be made clear exactly what is being compared." The reviewer cautioned that when the population is not well defined or the subgroups being compared are not clearly stated, the reader may apply the findings inappropriately.

Some reviewers expressed confusion about restriction to a very specific subset of the population vs. a more general subgroup analysis. One noted that specifying unusual eligibility criteria that other studies might have considered exclusion criteria would be important. The example was the AWESOME (Angina With Extremely Serious Operative Mortality Evaluation) trial of special high-risk patients with ischemic symptoms refractory to medical therapy who were at increased risk for adverse events after either percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG).

Another comment concerned the concept of applicability. The claim was that included trials are typically efficacy studies and that they do not provide information for real practice in which, for instance, patients have multiple comorbidities. Note: the draft AHRQ EPC guidance recommends distinguishing between efficacy and effectiveness studies and references work done by Gartlehner et al.[165]

Table 17. Types of comments received on draft comparative effectiveness reviews

Comment Types
Lack of information on clinically relevant subgroups
Unclear specification or definition of population studied
Randomized controlled trials forming foundation of research oriented around specific subgroups not relevant to "real world"
Failure to include all studies/data with relevant information on clinical heterogeneity
Inclusion of inappropriate studies/data (e.g., included study with special high-risk subgroup that was excluded by all other studies)
Inappropriate analysis or interpretation of subgroups data
Inadequate consideration/controlling for disease activity and severity
Problematic presentation of studies with heterogeneous populations
Conclusions fail to account for subgroups

Reviewers also pointed out cases in which studies or data with relevant information on clinical heterogeneity were never mentioned at all. They advised that if no studies existed that addressed important clinical subgroups, then EPC authors should state that fact clearly; if such data did exist but were not considered in a CER, then EPC authors should explain the reasons for excluding the studies or the data. For example, one reviewer noted that clinicians are very interested in patients with comorbid depression and chronic pain and this was not addressed in the report. Yet several studies were available that would shed light on this issue.

More generally, reviewers criticized the lack of consideration of disease activity and severity. They emphasized that these factors are very important for understanding drug efficacy and safety in patients with differing severity of disease.

Inappropriate analysis or interpretation (or both) of subgroup data were sometimes criticized. One critique focused on forced grouping of subgroups: "The mix of simple and complex randomized controlled trials in a forest plot with a summary line is simply inappropriate. There is a great hazard in blending trials with such divergent target populations as the authors have done in multiple forest plots." Whether this was intended to be a call for more discrete subset, subgroup, or heterogeneity analyses, however, was not clear.

Another point involved where in the report that subgroup analyses might be explored and discussed in more depth. Some comments advised that discussion or conclusion chapters of the reports should include commentary on subgroup analysis. For example, one commentator thought that this paragraph from the results section of the report should have been brought forth into the discussion or the conclusions section:

> We were interested in finding studies that would allow us to predict individual responses to a specific drug based on [a] patient's clinical and genetic characteristics. In theory, drugs have varying side effect profiles and [an] individual's tolerance of those side effects varies but overall incidence of side effects is relatively high. The lack of data relating individual's characteristics to drug effects makes it difficult to predict which drug will be best tolerated by a specific individual. This is indicated by substantial discontinuation rates and frequent need to try multiple drugs before finding an effective drug that is well-tolerated. Studies of tailoring therapy would have been eligible for this review, but we did not find any. Most of these studies looked only at average effectiveness, excluded subjects with comorbidities, and did not even assess difference in effectiveness according to broad demographic characteristics.

Despite the fact that the KQs for these three reviews did indicate the consideration of clinical heterogeneity, the reviewers of these reports critiqued them on which subgroups were evaluated, how the evaluation was done, and the lack of information on factors contributing to clinical heterogeneity.

KQ 4b. General Critiques (in the Literature) about Clinical Heterogeneity in Systematic Reviews

The most frequently mentioned concern noted for assessing clinical heterogeneity in SRs is that authors should specify in advance, during the development of the design or protocol for their review, which factors they will be investigating. Analysis of factors identified a posteriori may be considered a "data dredging" exercise that is likely to produce unreliable results.[4,5,11,38,166-185]

A related concern is that analyses identifying factors that appear to modify intervention-outcome associations should be regarded with caution. Factors investigated may not be biologically plausible or based on disease pathophysiology,[186] and may be misleading.[170] Subgroups arising from a "per protocol" rather than from an intention-to-treat analysis of randomized controlled trials may be particularly suspect because control of possible confounding by randomization no longer holds.[166,176,178-180,182,187,188] Hence, the analysis of potential indicators of clinical heterogeneity is considered hypothesis-generating.

Many authors suggest that requisite caution should be exercised by severely limiting the number of pre-specified factors and by controlling the overall type I error probability for the entire group of factors. The latter approach has the effect of reducing the type I error probability for each factor.[166,168,179,180,186,188-191]

Homogeneity tests have low power, and this problem can cause authors of SRs to miss clinical heterogeneity that has an impact on the magnitude of a treatment effect. For this reason, some experts suggest the use of a higher alpha level than usual, such as 0.10.[6,170,183,191,192]

Two approaches to understanding or dealing with clinical heterogeneity were popular in the earlier literature: excluding outlier studies without sufficient justification[11,178,182,193] and using the control groups of included studies to estimate the underlying risk of the outcome.[194-196] Given the risk of selection bias when excluding studies without just cause, this is no longer done. Currently, we are still using the rate of events in the comparison group to control for baseline risk.

KQ 5. What Evidence Is There To Support How Best To Address Clinical Heterogeneity in a Systematic Review?

This section describes the literature search to identify best practices on handling clinical heterogeneity in SRs and CERs and our discussions with key informants.

Review of Methodologic Studies Addressing Clinical Heterogeneity

As described in Chapter 2, we conducted a formal literature search in an effort to identify best practices for handling effect-measure modification within SRs and CERs (or constituent meta-analyses). The intent of this search was to identify whether guidance on the conduct of SRs and CERs (1) differentiated among different types of heterogeneity, and (2) described how to identify factors causing clinical heterogeneity, including evaluating particular subgroups and conducting analyses on individual patient data rather than using the summary results from publications.

For this question (which yielded more than 1,000 citations at the outset with an additional 387 identified via citation search), two senior reviewers independently reviewed the output and identified 60 publications that discussed how to handle clinical heterogeneity in SRs (broadly defined). After removing two duplicate articles, we initially had 58 papers for review.

We also reviewed an additional group of 25 papers that the SRC identified in its publication library that addressed clinical heterogeneity or subgroup analyses. The overall sample included 83 papers (summarized in Evidence Table C3 [Appendix C]). These 83 publications cannot be considered as representing either a systematic search or a random sample. Although the 1,000+ citations were identified from a formal systematic search, the final 83 papers reviewed were not independent with regard to authorship but rather exhibited extensive clustering by a small number of experts (Table 18).

Table 18. Clustering of authors of publications about clinical heterogeneity

Number of Papers With Same First Author	Last Name(s) of First Author(s)
6	Thompson
4	Ioannidis
3	Berlin, Schmid
2	Brookes, Chalmers, Feinstein, Higgins, Lau, Moher, Petitti

Of the 83 papers we reviewed, 80 (96 percent) addressed heterogeneity among studies in one form or another; the other three focused on evaluating heterogeneity in studies rather than in SRs. In all, 57 (69 percent) of the papers addressed within-study heterogeneity. There were 54 studies (65 percent) that addressed both within- and among-study heterogeneity, often in the context of comparing conventional meta-analysis with individual patient meta-analysis or distinguishing study-level study characteristics (e.g., randomized or observational) from patient-level characteristics (e.g., disease severity). These publications did indicate that analysis of individual patient-level data in meta-analyses does allow better assessment of clinical heterogeneity, but the time, cost, and difficulty in obtaining these data is often prohibitive.

Of these 83 studies, 53 (64 percent) distinguished between heterogeneity regarding methodologic characteristics of studies that would affect their internal validity (e.g., allocation concealment in trials) and characteristics of patients and clinical settings that would affect external validity (e.g., presence of coexisting conditions). The papers that did not draw this distinction tended to be those that focused more on the statistical aspects of heterogeneity assessment than on substantive applications.

Finally, 14 articles (17 percent) gave guidance for defining indicators or measures of clinical heterogeneity. For the most part, this guidance tended to be very general, such as using clinical judgment, conducting interviews with patients, and looking for leads in previous research.

At the current time, there is no guidance on how to identify which factors should be considered as potential effect-measure modifiers of the treatment-outcome association. The literature is very general and suggests the use of experts and information from the literature, but how does a systematic reviewer determine which literature is most relevant? For example, systematic reviewers typically include demographic factors such as age, sex, and race/ethnicity, with little forethought on why these factors may be relevant or if there are more critical effect modifiers that should be considered. Guidance and processes to determine how to select important effect modifiers is not available in the public domain currently.

Results of Key Informant Interviews

We interviewed six authors in all; three pertained to osteoarthritis reviews (one each from Cochrane, an author of a health technology assessment, and AHRQ via a search of the DARE database),[14,15,19] and three pertained to myocardial infarction (one from DARE, two from Cochrane).[16-18] RTI staff attempted to include authors of NICE SRs, but we were unable schedule to interviews because of the authors' limited availability within the specified time frame of this task.

Topical Analysis

Typical approach for developing a study protocol for a systematic review. Five of the six participants indicated that they follow a process or protocol, such as the process described in the Cochrane Collaboration guidance,[197] when developing a study protocol for an SR. Four

participants specifically mentioned use of the "PICO" scheme, which addresses the patient(s), intervention(s), comparison(s) (comparator[s]), and outcomes. This is a slightly abbreviated version of the "PICOTS" framework often used by EPCs, which also includes timeframe and setting elements. These observations generally pertain to the general process for developing a workplan or protocol for the entire review, not to particular elements such as stating KQs or outlining specific analytic techniques. Two participants noted that they also consult with experts in the field:

> "We would go through a preliminary literature search. We'll have our own conversations with experts that we have contacts with already," explained one participant. "We'll be in touch with whoever did the topic development in the SRC [Scientific Resource Center], but now that's being done more and more by the Evidence-based Practice Centers [EPCs] themselves. Whatever technical experts they consulted with, we try to make an effort to contact them…then it's all aimed at us gaining ownership of the topic to make sure that we have mastery of the issues as we are able to get started on the review."

Timing of subgroup identification and ideal process for subgroup identification. Participants were asked to indicate the specific point in the review process at which authors should formulate subgroups and the ideal process for identifying subgroups in an SR. Four authors said that subgroups should be developed during the protocol development process; however when asked specifically about a priori vs. a posteriori identification of subgroups, five participants indicated that subgroups should be identified a priori.

Timing of subgroup identification. With regard to the timing of subgroup selection, one author stated,

> "[P]eople should think about whether there are clearly defined, if you like obvious, subgroups of patients who may display or react differently to a given intervention. And if there is substantial evidence to support such an assumption, they should then plan appropriate subgroup analyses to investigate in their review if this is the case. Now this isn't to say that you're not allowed to do a subgroup analysis that you didn't pre-specify. I see research in general, but also systematic reviews and meta-analyses, as a creative process and exploration of the data is a good thing because you can discover something. But, of course, there is a risk that what you see is a chance finding if you have explored very extensively and you happen to fall upon a finding that's not real. For this reason, I think you should always make it clear whether a subgroup analysis was pre-specified or whether it was explored."
>
> Another commented:
>
> "I think, ideally, it should be done during the protocol development process and I think that's why we're so intent when we do systematic reviews to, as I said, gain ownership of the topic. To really become immersed as quickly as possible in the important issues because once the protocol has already been developed, I think there's a real interest in getting through the search and abstraction phases as quickly as possible, and it can be very disruptive to have to redesign your abstraction instruments midway through the process. It can be really frustrating

and there can be a lot of duplication so I would vote for the protocol development process."

A third noted:

"[I] guess it should be done early in the . . . protocol development. My experience has been that ideally it should be done almost in isolation, but the reality is that because of your previous knowledge of the literature, that's going to influence some of that subgroup development. Your knowledge may be a little biased on what you've already read or what you already know of the literature, so I guess early on in the development of the process or the protocol before you've even started your search."

A fourth said:

"If you can, it should be done before—a priori. That would be the best way of doing it, but sometimes, like I said, when you're reviewing literature some things stand out. You find a subgroup of patients dying more often than others and you probe a little farther. When you basically publish those results, they may not be as robust as a priori hypothesis, but still something may be clinically meaningful. The short answer is you should have a hypothesis before. If you want to look at subgroups, you should have a hypothesis generated before."

Along similar lines of reasoning, another responded:

"[I]f the question was: 'Is there meaningful difference between the subgroups based on age, race, or sex?' then we would have created a priori hypothesis and assessed the number needed to have meaningful differences in the subgroups. But, occasionally what happens is that when you do the analyses, some subgroups fall out (they look quite abnormal), and then we probe a little better into that but we do not make that our primary result of the analysis."

Two participants felt that clearly stating how and when authors specified subgroups is of great importance.

"[T]here's nothing wrong with doing subgroup analyses that were not pre-specified. There may be some new finding that wasn't known at the time when you wrote the protocol that leads you to think about subgroup analyses that you didn't think about before … you could just creatively explore the data, but in this case, you should make it clear that this is how it happened and that it wasn't pre-specified."

"What I do and what I recommend as an editor at the Cochrane Collaboration, we constantly say that if people tend to, or plan to, interpret the statistical analysis in terms of inference afterwards, they are supposed to present in the protocol what key clinical characteristics they would consider relevant, or in terms of exploring it afterwards. So the main issue here, which I'm feeling very strongly about, would be if they were supposed to interpret it. Actually, I did a meta-analysis on weight loss for knee osteoarthritis, and in that, we had a strong a priori saying we wanted to include the dosage that would be the average weight lost as a covariate. That was an example of something we knew prior to doing the statistical analysis. But in the [other] paper, we did it the other way around; we wanted to explore

reasons for heterogeneity. So if we used a statistical analysis package to generate an inference, then it should be carefully stated in the full paper. That's a very strong argument and I really feel strongly about that because it's obvious that very often we see people doing whatever subgroup analysis or meta-regression analysis and sometimes it seems that they have been inspired to do that following looking at the data."

One author suggested a two-step process for identifying subgroups that entails (1) looking a priori for patient populations that make clinical sense to a clinician working at the bedside and to opinion of clinical leaders as well—looking at the knowledge that has already been developed; and (2) looking after the fact at heterogeneity that is evident in the results to see if that can highlight patient characteristics or study characteristics that lead the development of subgroups.

***Ideal process for subgroup identification for studying clinical conditions in an SR*. The** participants provided a range of responses when asked about the ideal process for identifying subgroups in an SR. To identify subgroups, one author said:

"Authors should, together with content experts, consider what's clinically relevant … in terms of what we anticipate would mean something to the response to therapy. That should be based solely on external knowledge without looking at the data. It's very important that the content expert is not involved in the data handling and that the ideas for how you are supposed to explore reasons for heterogeneity is made a priori. That would mean that the protocol is based on content expertise rather than looking at the particular study."

Two participants cited the importance of considering additional sources of potential information, such as content experts or a literature review, when attempting to identify subgroups. One felt that identification of subgroups should include:

"A blend of leaning on the usual suspects like age, disease duration, disease severity, sex—those kinds of things that are almost always considered—and then also leaning on what the literature suggests might be important subgroups. That's why it's so important before even beginning on the review to have done a preliminary literature search, to get a sense of what important subgroups might exist, and also talking with experts who may already be familiar with what subgroups exist."

Two participants noted that subgroup analyses are feasible when they have sufficient numbers of studies. One participant noted that many Cochrane reviews include too few studies to do a subgroup analysis and the other participant added that to do subgroup analyses that combine data across studies, patient demographics, treatments or interventions, and outcomes should be fairly homogeneous. He further elaborated saying,

"[F]or my particular study, there was little information at the time of the study, so we had a small number of patients . . . .[T]he four randomized trials that we eventually identified to include in our analyses were fairly small and we had a total of only seven to twenty-five patients. So therefore, that itself limited us to do subgroup analyses for our studies. As you know from our study, we really didn't

do any subgroup analysis . . . we did not stratify the results based on subgroups because the numbers were too small."

Considerations when developing key questions. Four of six participants considered demographic factors when developing the KQs for their SRs. One participant acknowledged that all the included factors were data-driven and were considered post-hoc because the authors did not anticipate that they would be able to find many relevant studies. He felt that all studies should be pooled and split again afterwards to reduce differences in clinical heterogeneity. One participant did not consider demographic factors at all, and another stated that there were too few relevant studies for subgroup analysis.

Half of the participants considered disease severity during development of the KQs. One participant noted that he considered disease severity post-hoc. Another said that severity will be a factor that will be considered in the future but not at present, given the limited number of studies available pertaining to the topic. One author indicated that the author team did not consistently look for severity across the studies.

None of the three authors of osteoarthritis SRs considered the affected joint when developing the KQs for their review. For example, one said,

> "The particular project that we did was focused on OA [osteoarthritis] of the knee, so that wasn't a specific issue and whether it was the right or left knee wasn't really a concern to us."

Three participants considered disease recurrence, one did so post-hoc, and the remaining two did not account for it at all. However, one author of a myocardial infarction review qualified his response:

> "If the patient didn't have troponin* or ECG [electrocardiogram] changes, then they had to have chest pain in the setting of a previous MI [myocardial infarction]. When we developed the questions, we were actually thinking of it that way so we did kind of include that, but it was by chance, rather than by design."

Consideration of other clinical factors. Participants mentioned several other clinical factors that they considered when developing the KQs for their reviews. Among them were timing factors such as time between symptom onset and intervention administration, duration of trial (i.e., sufficient time for effectiveness to be noted / measured); prior or concurrent interventions; baseline risk factors (i.e., body mass index); and whether the disease was classified as primary or secondary (i.e., the reason for the trial vs. a comorbidity in trials where other conditions were primary). One participant stated,

> "[C]linically we were interested in looking at the early phase of treatment. So we were trying to limit our inclusion criteria to location of treatment thinking that would be a proxy marker for 'acutes.' So we were trying to look at patients that were included only very early on in their presentation. . . . [W]e also struggled with looking at outcomes to look at how far out should be an appropriate look at outcomes."

* The troponin test is used to help diagnose a heart attack, to detect and evaluate mild to severe heart injury, and to distinguish chest pain that may be due to other causes. In patients who experience heart-related chest pain, discomfort, or other symptoms and do not seek medical attention for a day or more, the troponin test will still be positive if the symptoms are due to heart damage. Troponin tests are often preferred as they are more specific for heart injury than other tests (which may become positive in skeletal muscle injury) and remain elevated longer. (http://www.labtestsonline.org/understanding/analytes/troponin/test.html).

During our recruitment efforts, we contacted one author who declined to participate, indicating that her SR pertained to exercise after knee arthroplasty for osteoarthritis rather than osteoarthritis and exercise. However, her email offered insight on factors that her team considered when developing the KQs for their SR. Regarding diagnosis and disease severity, the patients were post-arthroplasty (i.e., all patients were considered to have sufficiently severe disease that had not responded to all previous forms of treatment and thus warranted total knee arthroplasty). All had clinical and radiographic changes that led their providers to perform the operation. As part of their inclusion criteria, the patients had to be able to undertake an exercise rehabilitation intervention. Thus, the authors of this review did consider clinical heterogeneity: both severity of disease and baseline risk factors for carrying out the planned intervention.

Selection of factors to report in the analysis. Two authors indicated that they selected factors a priori to include in their analysis. One author elaborated on the importance of having a clinical hypothesis, saying,

> "[W]hen I start extracting data, I have a very optimistic outlook, so to speak. I try extracting almost whatever and if I don't feel strongly about any specific covariates—those that we are discussing here, study characteristics, clinical heterogeneity reasons . . . when writing the protocol, if I allow one cell to be blank then I obviously don't feel strongly about it. If I'm feeling like I have too many blanks, like 50 percent blanks, I omit it from the publications table. If that's the case then obviously I don't feel that strongly about it and I only include it in terms of external validity, making sure that the readers of the full paper feel confident that they understand the paper. But the thing about statistical analysis … if I'm supposed to believe in my own results afterwards, I need to have a clinical hypothesis."

One participant mentioned that the small number of studies included in his team's SR precluded them from even looking at clinical heterogeneity. Another noted,

> "[O]ur stratification or our subgroup analyses were driven by the discussion around the discrepancy between the small studies showing substantial benefit and the very large ISIS-4 study [Fourth International Study of Infarct Survival] that didn't show any benefit at all. . . . I'm not sure that this is a very typical situation because it was really driven by this very large study which didn't show any effect and discussion about if it was possible how the smaller studies had shown quite substantial benefit and these results were then nullified, if you like [and] were not confirmed by the large ISIS-IV trial."

Another participant indicated that, for his team's SR for myocardial infarction, they referred back to evidence from previous SRs and correspondence with authors to decide whether something might be an important contributor to clinical heterogeneity. The authors considered doing a subgroup analysis looking at patients who were troponin positive. Given that these patients would be potentially sicker or have more severe disease than others who were not troponin positive, the authors thought that these patients might respond differently than patients who were troponin negative. They also hoped to do a similar subgroup analysis looking at patients with positive electrocardiogram changes, reasoning that such changes may indicate a different disease or different severities of disease. However, the data were not available to conduct such analyses in either case.

One author noted the importance of using tests of interaction to evaluate the strength of the evidence for the differences between subgroups.

> "[Y]ou should always use appropriate statistical tests to investigate to what extent differences you observe between two subgroups are real or mainly the play of chance."

Reference to guidelines and manuals during study protocol development. All six participants mentioned use of or reference to the Cochrane manual during study protocol development; three of the participants were, of course, authors of Cochrane reviews. Other guidance documents mentioned included QUORUM,[198] PRISMA,[199] the paper by Harris et al.,[200] and the RTI-UNC EPC report by West et al. pertaining to rating the strength of evidence of different systems.[201] The participants used the documents as reference tools to help resolve problems that they may have encountered during study protocol development, and also to assist with quality assessment of the included studies.

Additional considerations in the selection of patient or disease factors. Several authors provided additional thoughts on determining which factors to consider for assessing clinical heterogeneity.

Limited number of relevant studies published. Although he had mentioned it earlier in the interview, one participant thought it important to reiterate the limitation of having a small number of relevant published studies when conducting an SR.

Benefits of the PICO format. Many of the participants valued the PICO format. One participant stated:

> "I'm very happy to recommend that people use the PICO format, as that's how meta-analyses are supposed to be written. When authors use the PICO [format], then we know that all the studies that they are able to consider eligible should be pooled per se and then the I^2 [inconsistency index] is very important. . . . If you have lots of very different studies, but they all fulfill your PICO framework initially a priori, then they should be pooled. But if the I^2 goes nuts, meaning that it's far too high, say extremely over 50 percent, then the overall estimate is not relevant. Then you need to explore in more detail why the I^2 went nuts . . . [M]y overall conclusion would be that we should always combine the studies that fulfill the PICO framework that they considered initially, but if the I^2 goes nuts, they should not put that much emphasis on the overall results and make sure to continue and explore reasons for clinical heterogeneity."

Exclusion of poor-quality studies. Quality of studies should be evaluated before inclusion in an SR. One participant remarked:

> "[W]e made a great effort to really identify all studies. We went into the Chinese literature, and there was even some hand searching in the Chinese literature. All we found was all positive and not very well conducted studies. I personally think the problem here wasn't publication bias, it was just low-quality, inadequate quality, bad, small studies."

Scope of inclusion/exclusion criteria. Development of the study questions largely influences the extent to which clinical heterogeneity is addressed. As noted by one participant,

> "[W]hen we were developing the protocol, our study question was very specific in that we were looking for patients initially who presented to the emergency departments with acute coronary syndromes [ACS]. That was our initial question and when we submitted that to the Cochrane [group], if I remember correctly, the review group came back and wanted us to make this a broader review so that we would include inpatients in the analysis. So we had a pretty lengthy discussion and one of the issues that we felt was, in part, around heterogeneity is that inpatients that developed ACS were in fact different from those presenting [to the] emergency department with ACS. We successfully were able to argue or communicate our point to the review group, so we kept things fairly narrow, but our biggest tool for dealing with clinical heterogeneity was initially in developing the question and that took a fair amount of revisions to make sure that we had fairly narrow definitions of what we would include and what we would not include."

Summary

- This report focuses on clinical heterogeneity, which we define as the variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in an SR or CER that may influence or modify the magnitude of the intervention measure of effect. This is distinct from methodological heterogeneity, which refers to variation in study designs and analyses as reasons for differences in treatment effects among studies.
- All five organizations (AHRQ, CRD, Cochrane, DERP, and EUnetHTA) refer to variation in population characteristics, interventions, and outcomes to define clinical heterogeneity. AHRQ, Cochrane, and CRD use the term "clinical diversity" rather than "clinical heterogeneity."
- The underlying rationale of statistical tests to assess heterogeneity is to investigate whether existing variations in treatment effects go beyond what would be expected by chance fluctuations alone. Commonly used statistical tests are Cochran's Q test, I^2 index, and meta-regression.
- Common reasons for statistical heterogeneity include clinical heterogeneity, methodological heterogeneity, chance, and biases. False conclusions about clinical heterogeneity based on statistical heterogeneity result in false-positive conclusions (type I error) or false-negative conclusions (type II error).
- Generally, reviewers use one or more of three common approaches to explore heterogeneity: stratified analyses of homogenous subgroups, meta-regression, and sensitivity analyses. We also consider restriction as a way to understand clinical heterogeneity.
- We did not find guidance documents, studies, or commentaries indicating that clinical heterogeneity should be considered at all stages of the review.
- Most EPC authors considered demographic variables such as age, sex, race, and ethnicity, and variables reflecting coexisting disease in their subgroup analyses.

- Key informant interview respondents generally agreed that subgroups should be developed during the protocol development phase (a priori); several consult with experts in the field during the process and recommend this as a best practice. They tended to rely upon Cochrane Collaboration guidance and the PICO(TS) scheme in their review processes; some also referred to QUORUM, PRISMA, and methodological papers.
- Similar to studies reviewed for this report, key informant interview respondents tended to consider disease severity, disease recurrence, and demographic factors in assessing clinical heterogeneity.
- Studies assessing clinical heterogeneity methodology often conclude that systematic reviewers include demographic factors with little forethought about why these factors may be relevant or whether they should consider other, possibly more critical factors. However, guidance and processes to determine how to select important potential effect-measure modifiers is not readily available.
- Analysis of individual, patient-level data in meta-analyses allows for better assessment of both within- and across-study clinical heterogeneity, but the time, cost, and difficulty in obtaining these data are often prohibitive barriers to following such practices/procedures.

Chapter 4. Discussion

This chapter discusses our findings for the five key questions (KQs) posed by the Agency for Healthcare Research and Quality (AHRQ) for this methods report and presents our suggestions for further work related to addressing clinical heterogeneity in SRs, CERs, and meta-analyses. We discuss the definitions of clinical and statistical heterogeneity and how different review groups, including AHRQ, handle clinical heterogeneity in the conduct of their reviews. Also, we provide a summary of best practices as identified by the published literature and via key informant interviews of the authors of SRs we reviewed as part of this project. We note the limitations of both this review and the larger body of knowledge on which it is based and, finally, make some suggestions for further work.

The focus of this report is on clinical heterogeneity but our first KQ required us to compare clinical and statistical heterogeneity. In doing this comparison, we determined that the literature addresses methodological heterogeneity as well, which includes issues related to study design, study conduct, publication bias, and funding issues. An example of a methodological issue is whether the randomized controlled trial masked those who were to assess the outcome of the trial. In this review, we keep distinct factors that describe the individual patient (clinical heterogeneity) from factors that can be regarded as study design or methodological issues. When appropriate, we have contrasted clinical and statistical heterogeneity with methodologic heterogeneity, with a more formal comparison provided for KQ 1.

The remainder of this section recaps the KQs in turn and provides our interpretation of some of our findings. A later section comments on the limitations of the project, and a final section lays out some research issues that further work, which we were unable to pursue at this time, might help illuminate.

The Concept of Clinical Heterogeneity

The term "heterogeneity" as used in the epidemiology literature and assessed in clinical studies refers to an intervention-disease association that differs according to the level of a factor under investigation. The term "effect-measure modification" is often used to clarify that heterogeneity can be observed on the relative scale, the absolute scale, neither, or both, and may be present on one scale but not the other (hence, it is the specific *effect measure* where the heterogeneity is observed).

The presence of effect-measure modification may suggest a biologic (or etiologic) effect of a factor upon the intervention-disease relationship, or it may reflect one or more biases. A factor can modify an effect measure for the intervention-disease relationship when baseline rates of the disease vary among factor subgroups or when the baseline rates do not vary among those subgroups. However, it is important to note that baseline rates may vary within subgroups of a factor whether or effect-measure modification is observed on any scale. This is because whether a given factor modifies baseline risk of disease and whether or not it modifies the effect of a particular treatment on that disease, or the direction or degree to which it modifies that treatment effect, are unrelated. Our focus here is on *clinical* heterogeneity, which is not to be equated with effect-measure modification or statistical heterogeneity.

Many different clinical factors can be evaluated as influencing the intervention-disease association, including demographics (age, sex, and race/ethnicity), severity of disease, coexisting conditions, and cointerventions. For example when sex can be shown to be effect-measure

modifier, the estimate of effect (such as an odds ratio or risk ratio) is different for males and females. That effect-measure modification exists can be seen visually using contingency table analyses that are stratified by the factor, but often the importance of differing effect estimates is quantified using heterogeneity tests in stratified analyses or statistical interactions when modeling.

Evaluation of effect-measure modification of an intervention-outcome association is very important to a variety of audiences: those doing SRs or CERs (hereafter combined into a single category of SRs); practicing clinicians who need to understand the results of such reviews; patients; and policymakers at many levels, who must make numerous decisions about services to offer or cover in health plans, insurance schemes, and other programs.

Determining which factors influence the intervention effect helps identify which patients will likely benefit most, the characteristics of those who will likely receive no benefit, and perhaps more importantly, who might be at greatest risk for adverse events. If clinicians and others are provided with information that allows them to tailor treatment most effectively, they should be able to improve the health outcomes of their patients.

Thus, we examine here what clinical heterogeneity means to different groups and how it differs from (or complements) notions of statistical or methodological heterogeneity. Later KQs address—in the context of clinical heterogeneity— what effect-measure modifiers seem to be commonly used to address clinical heterogeneity and, when they are examined in materials such as SRs, where they first arise as matters of concern.

Clinical Heterogeneity Definitions by Different Review Groups

Definitions. In our comparison of how different SR groups define clinical heterogeneity, we recorded almost identical definitions from several important international organizations producing or tracking SRs; these groups include AHRQ,[22] the Centre for Reviews and Dissemination (CRD),[25] the Cochrane Collaboration,[20] the Drug Effectiveness Review Project (DERP),[23] and the European Network for Health Technology Assessment (EUnetHTA).[29] All use a definition framed as variability in the populations studied, the interventions involved, and the outcomes measured. These are three of the six factors that reviewers supported by AHRQ's Evidence-based Practice Center (EPC) program considered in the development of KQs for AHRQ SRs (PICOTS, which in its most expansive format refers to populations, interventions, comparators, outcomes, timeframes, and settings).[22]

The main difference among these definitions is the terms they use to cover the general concept. For example, rather than specifically using "clinical heterogeneity," AHRQ, Cochrane, and CRD use the phrase "clinical diversity." DERP, by contrast, uses "qualitative heterogeneity" in its manual. Nevertheless, we conclude that consensus around a common definition and explication of associated terminology would be valuable to the field of SRs.

Restriction. Our review also encountered the idea of restriction—limiting the studies in a review to those whose participants reflect a clearly restricted subgroup—as a way to address clinical heterogeneity. We acknowledge that some may view the notion of restriction as a way to avoid dealing with clinical heterogeneity altogether, rather than as a way to address variability in factors such as patient characteristics or outcomes. Of the manuals we reviewed, Cochrane and CRD alone mentioned restriction explicitly. Both caution against this technique for addressing clinical heterogeneity because it limits the applicability (i.e., generalizability) of results reported in SRs.

Regarding restriction, however, we reasoned that reviews of studies that focused on a specific, probably narrowly defined subgroup, can provide valuable information for tailoring care for individuals with the same characteristics as in that subgroup, despite their inability to provide information relevant to individuals outside that subgroup. For that reason, we do not necessarily caution against use of this approach. Whether restriction should be part of the toolkit for addressing clinical heterogeneity is an issue that needs to be considered more broadly.

Clinical Heterogeneity vs. Statistical Heterogeneity

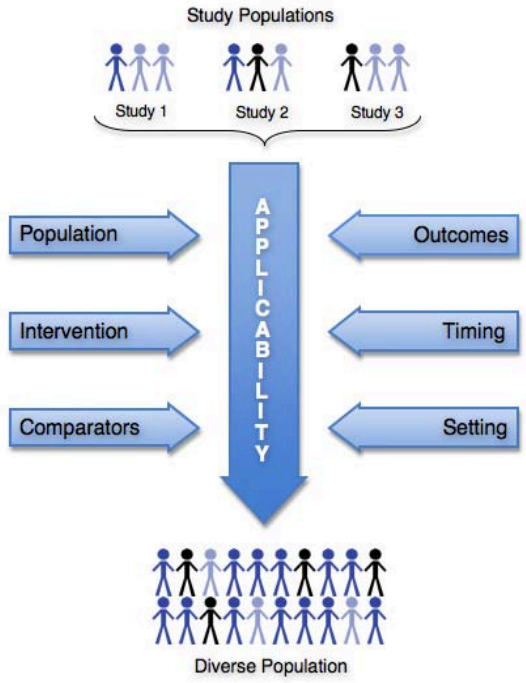
Clinical heterogeneity is closely linked to statistical heterogeneity, especially because the occurrence of clinical heterogeneity may lead to statistical heterogeneity[6] detected using techniques such as Cochran's Q test,[34] the I^2 index,[35] or meta-regression.[10] Statistical heterogeneity is hypothesis driven; the null hypothesis is essentially that the effect of treatment does not differ by levels of factors such as age, sex, disease severity, settings, or study quality, given an acceptable probability of type I error. Clinical heterogeneity has the same underlying hypothesis but the concept does not include study design (methodological) issues and is not framed in such mathematical terms. In addition, statistical heterogeneity is broader than clinical heterogeneity in that it encompasses clinical and methodological heterogeneity as well as chance (random error).

For both clinical and statistical heterogeneity evaluation, the factors that may cause heterogeneity should be identified a priori, during the protocol development stage. Across all the elements of this study, the question of when to specify effect-measure modifiers to examine in SRs to address clinical heterogeneity was a matter of some contention. Overall, however, the trend seemed to favor a preference for a priori identification of factors that could lead to clinical heterogeneity. When a priori identification does not occur—that is, when analyses are done on factors or subgroups determined post hoc—some experts refer to this as "data dredging"[202] and generally caution against it. Whether broader consensus can be reached on the desirability of identifying subgroups or effect-measure modifiers for analysis only a priori remains an open question.

Clinical Heterogeneity vs. Applicability

Clinical heterogeneity is closely related to the idea of applicability, and we show this relation in Figure 4, which includes factors that contribute to clinical heterogeneity (PICOTS). When assessed, this concept (often referred to as external validity or generalizability) tells readers whether they can infer that the intervention-outcome association studied can be broadly generalized to different populations, varying treatments, outcomes, or settings. More formally, it (as external validity) has been defined as "Inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments, and outcomes."[41]

Figure 4. Relation between clinical heterogeneity and applicability in systematic reviews



Whereas clinical heterogeneity reflects the presence and extent of differences in population characteristics and coexisting conditions and treatments within the set of studies included in an SR, applicability takes such findings to the next level (i.e., translating the information on clinical heterogeneity into clinical practice). In other words, the concerns extend to the degree to which the findings in SRs are directly relevant to broad patient populations in a wide range of settings. The extent to which SRs deal broadly and explicitly with effect-measure modifiers in the five categories noted earlier (demographics, disease variables, etc.) puts a ceiling on how expansively findings can be said to apply to patient subgroups of interest to clinicians or policymakers.

Clinical Heterogeneity in Systematic Review Key Questions

To address this question, we evaluated how SRs produced by four review groups and those identified in two CRD databases (Database of Abstracts of Reviews of Effects [DARE] and Health Technology Assessment [HTA]) addressed clinical heterogeneity when the authors developed their KQs for 15 clinical conditions: breast cancer, lung cancer, prostate cancer, cesarean section, chronic kidney disease, chronic obstructive pulmonary disease (COPD), depression, dyspepsia, heart failure (including congestive heart failure), heavy menstrual bleeding, hypertension, irritable bowel syndrome (IBS), labor induction, myocardial infarction, and osteoarthritis. Not all sources had SRs for all conditions, but across all groups and conditions we identified in all more than 120 reviews to use in this analysis.

We focused on whether the reviewer groups considered variables in five separate categories: (1) demographic characteristics of patients and subjects, such as age, sex, race, or ethnicity, or less commonly factors such as insurance coverage or income or socioeconomic status; (2) disease variables including stage, type, and severity; (3) risk factors for the principal

condition in question; (4) cointerventions that might have been used in conjunction with the therapy under consideration; and (5) coexisting conditions.

Key Questions in Reviews by Specific Review Groups

AHRQ. We reviewed 11 AHRQ reviews produced by various EPCs (for all conditions except chronic kidney disease, heavy menstrual bleeding, IBS, and myocardial infarction). With the exception of the review on COPD,[51] all the reviews specified factors for addressing clinical heterogeneity in their KQs.

Demographic factors such as age, sex, and/or race or ethnicity were typically included;[46,48-50,52-54,56] both coexisting conditions[47-50,52,54,56] and cointerventions[47,52-54,56] were often considered. Disease variables, such as stage or severity, or risk factors were less often specified in advance.

The AHRQ topic nomination and refinement processes are generally separate from an EPC's production of the actual SR; thus, specification of factors of interest with respect to clinical heterogeneity may not rest with EPC authors but rather with organizations or societies that suggest topics in the first place. Given AHRQ's process for formulating KQs (or subquestions), variation in this aspect of KQs may be driven as much or more by the topic development process than by the methods that EPC reviewers use. Nevertheless, EPC reviewers can always propose to give further attention to subgroups or to conduct analyses by important effect-measure modifiers and variables; the critical point is that proposing such analyses ideally should be done at the outset of the review, not after the fact.

No AHRQ SRs specified in detail how they arrived at factors to include in their analyses. This omission likely reflects the processes of the funding agency rather than specific methods or interests of the review group.

Cochrane Collaboration. In contrast with AHRQ, only five of the nearly 40 Cochrane reviews for 14 clinical conditions (all except dyspepsia) included demographic factors as part of their KQs.[64,82-84,87] Several reviews also included consideration of disease variables.[14,64,67,72,83,84,88,91,92]

Of the 14 conditions for which we had reviews, hypertension was the one clinical condition in which heterogeneity seemed to be considered in the development of KQs and reflected in the review itself. We speculate that the extensive literature base for this condition may allow detailed consideration of clinical heterogeneity. Whether the review groups had "prespecified" which factors to consider for assessing clinical heterogeneity during their initial protocol development stage, however, is not clear from the final reviews.

Reviews from the DARE database. We identified 37 SRs from the DARE database (none for dyspepsia, heavy menstrual bleeding, or labor induction).Three reviews included a demographic factor in their KQs (age,[111,113] race[119]); 10 included a disease variable[95,96,105,106,108,110,117,125-127] and 3 considered risk factors.[95,102,122] Population subgroups were not pre-identified or there was no information on whether and how they were identified. We found no trends with regard to clinical condition because so few reviews identified from the DARE database addressed clinical heterogeneity in the phrasing of their KQs.

DERP. Reviews conducted for this project are focused on pharmaceuticals rather than conditions. Many drugs can be used to treat several conditions; as a result, the same DERP reports may address multiple conditions (see Table 12). We reviewed 18 reports for eight

medical conditions (six duplicates; no DERP SRs were available for breast cancer, lung cancer, prostate cancer, cesarean section, chronic kidney disease, heavy menstrual bleeding, or labor induction). Of the 12 unduplicated reports, 10 specified the demographics they included[128-133,135-138] and all 12 considered disease variables, risk factors, cointerventions, or coexisting conditions.[128-139]

This consistent treatment of clinical heterogeneity can be explained by the fact that the last KQ in DERP reviews typically refers to issues related to clinical heterogeneity. All the factors considered in the review had been laid out in the initial KQs. However, none of the reviews gave specific information on how the subgroups had been identified to begin with. DERP reports go through a fairly complex development phase before the reviews commence, and the sponsor base (chiefly US state Medicaid agencies) has clear (and known) interests in the patient populations whom they serve. Thus, some effect-measure modifiers are likely to be commonly expected for attention in DERP reports at the outset.

Reviews from the HTA database. We obtained SRs for eight medical conditions: breast cancer, lung cancer, prostate cancer, depression, dyspepsia, hypertension, myocardial infarction, and osteoarthritis[140-149] (none for cesarean section, chronic kidney disease, COPD, heart failure, heavy menstrual bleeding, IBS, or labor induction). Of the 11 reports from this source, three addressed one or another of these heterogeneity categories:[19,140,144] two demographic factors[19,140] and three either disease variables or risk factors or both.[19,140,144] One of these reviews had been carried out by an AHRQ EPC.[19]

Unlike SR groups discussed above, one set of authors of a technology assessment review identified through the HTA database documented how they determined which population subgroups to review.[142] Specifically, they based their decision on information from work done by Gail et al.[155] on risk factors for breast cancer and Parmigiani et al.[156] on the BRCAPRO for BRCA1/2 mutations.

National Institute of Health and Clinical Excellence (NICE). We reviewed five NICE reports: breast cancer,[120] depression,[116] dyspepsia,[117] myocardial infarction,[118] and osteoarthritis.[119] Two reports specified demographic variables and clinical variables of one sort or another in their KQs;[116,118] two others provided information on clinical factors;[117,120] the one remaining NICE review[119] did not specify any factors that might contribute to clinical heterogeneity. However, this review[119] and two of the others[116,118] did provide information on how they arrived at which factors to evaluate in their review.

Best Practices for Developing Key Questions

Overall, the SRs carried out by researchers involved in the AHRQ EPCs, which includes those conducting DERP reviews, have tended to consider factors that may contribute to clinical heterogeneity as part of their KQ development. The same appears to be true for the NICE reports but we reviewed only a small number (but with many different KQs) to gather this information. The reviews we identified from the Cochrane and CRD databases tended not to specify potential effect-measure modifiers of the intervention-outcome association as part of their KQs. However, we do know that Cochrane and CRD reviewers, according to their manuals reviewed with respect to definitional issues, do clearly define clinical heterogeneity (as clinical diversity) and suggest that the possible effect-measure modifiers contributing to clinical heterogeneity be identified during the protocol development phase.

The major concern with post hoc identification of effect-measure modifiers was discussed for KQ 4b, namely that identifying factors post hoc has been termed "data dredging" and, therefore, at risk of producing spurious results. Also, concern has been raised about controlling for type I error when many potential subgroups are evaluated for effect-measure modification.

Other than guidance on when to identify possible effect-measure modifiers (a priori), we did not find very much information on how authors of SRs either have or should identify these factors. The AHRQ *EPC Methods Guide* suggested clinical experts or a preliminary literature review but the other manuals did not provide operational guidance. Of the SRs we abstracted for Evidence Table C2, only the NICE reviews appeared to use experts or the literature to guide the development of their KQs with regard to clinical heterogeneity. Review of Evidence Table C3, which summarized methodologic publications regarding clinical heterogeneity also shows that only 14 articles (17 percent) provide any guidance on identifying effect-measure modifiers either a priori or post hoc.

Consultation with clinical experts to identify which factors to assess for potential effect-measure modification is valuable for several reasons. First, they are typically very knowledgeable about the condition under study, knowing how effective current treatments are as well as potential adverse effects. As a result, clinical experts can help frame the KQs that clinicians need answered to best treat their patients. Clinical experts not only treat average patients, but typically see the atypical patient whose clinical factors may provide insight for the average patient. Finally, clinical experts are in touch with other experts and involved in their professional societies so they can better inform current practice gaps.

Based on the information that we derived from our review of guidance manuals and the methods papers we identified for KQ 4, we concur with the existing guidance to use clinical experts and/or a brief literature review to identify factors that might contribute to clinical heterogeneity and that these factors should be identified a priori. Our six key informants (KQ 5), agree with identifying factors a priori but stress that systematic reviewers need to disclose how the effect-measure modifiers were chosen and whether they were identified a priori or post hoc in all SRs.

How Systematic Reviews Dealt With Clinical Heterogeneity in the Review Process

Review of Guidance Documents

Some of the guidance manuals reviewed for information about definitions of clinical heterogeneity do discuss biological or clinical factors that may reflect clinical heterogeneity in one or more of the five categories noted earlier.[20,22,25,203] They agree that authors of SRs ought to consider them during the protocol development phase of the review (i.e., a priori).

With regard to handling clinical heterogeneity during the analysis phase, the AHRQ *EPC Methods Guide*[22] suggests that authors describe how they identified the factors they considered and how they determined whether pooling was or was not appropriate. The manual is not prescriptive on how heterogeneity should be handled but it states that when EPCs employ meta-analysis, they should conduct sensitivity analyses as well.

The Cochrane manual[20] and the CRD manual[25] provide more specific guidance on handling heterogeneity. They suggest that authors use visual exploration of heterogeneity with

forest plots and evaluate confidence limit overlap and that, if clinical heterogeneity is apparent, they follow up with subgroup analysis and meta-regression.

Subgroup analyses are acceptable according to AHRQ, CRD, and Cochrane guidances.[20,22,25] The Cochrane handbook stresses the need for transparency in the selection of subgroups for analysis. As noted earlier, however, it does not recommend restricting to narrow subgroups.

The Cochrane manual is much attuned to issues of applicability (i.e., external validity or generalizability, as discussed earlier). We did not focus on applicability in this report but we acknowledge that AHRQ is providing guidance on assessing applicability in its manual, and a recent paper on grading the evidence discusses applicability to some extent.[33]

Review of AHRQ Systematic Reviews

To see how AHRQ SRs had dealt with clinical heterogeneity, we abstracted information from AHRQ SRs for 11 conditions.[46-56] Our limited selection is not a comprehensive set representing all reports completed by the AHRQ EPCs for these conditions over the years. No AHRQ SR was available for chronic kidney disease, heavy menstrual bleeding, IBS, or myocardial infarction.

To determine whether the authors considered demographic and clinical values during the analysis phase, we focused on the results section of each review. Except for lung cancer,[47] these AHRQ reviews considered one or more demographic variables in their analysis.[46,48-56] All included at least one clinical factor from among the other four categories (i.e., disease variable, risk factor, coexisting condition, or cointervention).

Demographic characteristics. Contrasting factors specified in KQs with those addressed in the analysis (Table 9 vs. Table 15 in Chapter 4), we saw, not surprisingly, appreciable overlap for demographic factors that these AHRQ reviews considered and then either analyzed or tried to analyze, depending on what the literature base permitted. By and large, EPC reviewers will examine all subgroups defined by demographic variables that can be found in the literature base. For both COPD[51] and labor induction,[55] the authors did not identify effect-measure modifiers in their KQs; however, age was a modifier for labor induction and race/ethnicity and sex were evaluated for COPD. The authors of the lung cancer review did not plan (according to their KQs) to evaluate whether effect-measure modification arose from any demographic factors; neither did they evaluate demographic factors in the analysis.[47]

Generally, we conclude that if KQs demand attention to demographic variables, reviewers will be able to satisfy this requirement reasonably well. This statement is truer for standard demographic characteristics, such as age and sex, than for variables less commonly recorded in empirical studies, such as income, education levels, or language. Race and ethnicity may also be problematic insofar as these data may not be well specified in studies; for example, these variables may be rendered only as white vs. nonwhite. In addition, clinicians and others increasingly recognize that "culture" or "cultural background" may be meaningfully different from a simple race or ethnicity designation, and these constructs may not be measured at all. Finally, issues of language spoken in the home (or first language) and literacy or health literacy (in English or in first language) are also seen as of greater importance than thought in past decades, and these demographic variables are also rarely measured in clinical trials or studies.

Clinical variables. With regard to disease variables, risk factors, coexisting factors, and cointerventions, 4 of the 11 AHRQ reviews had specified plans to evaluate the same type of factors as they actually analyzed as part of the review.[48,49,54,56] For the remaining reviews, the authors either initially considered more clinical factors than they eventually analyzed[47,53] or, conversely, originally planned to analyze few clinical factors but ended up analyzing more than initially considered.[46,51,52,55]Also, one review considered risk factors and cointerventions at the outset but analyzed disease variables instead.

The discrepancies have three likely explanations. In the first instance, that fewer clinical variables ended up in analyses than had been originally planned is likely a reflection of the extent of information in the available literature. This is not an uncommon problem. Stakeholders and EPC analysts want information from SRs to be as pertinent as possible to clinical populations, but available studies (or at least those that can be regarded as of good or fair quality) simply may not address the desirable range of subgroups that can be defined according to these variables. Second, any added complexity involving demographic and clinical effect-measure modifiers together reduces the likelihood that such complicated (and narrow) subgroups will be adequately covered in included studies. Modeling might help overcome this challenge, but using such techniques are uncommon and open to their own criticisms. Third, in some cases evidence is quite compelling about a subgroup that had not been predicted ahead of time. Reviewers may believe, perhaps for persuasive reasoning based on biologic plausibility, that useful information will come from more in-depth examination of data on such subgroups; they may proceed to do these analyses even if such work had not been planned ahead of time.

Best Practices for Addressing Clinical Heterogeneity in Analyses

Somewhat in parallel with earlier questions about best practices in specifying clinical heterogeneity factors in KQs, another issue raised for this project targeted best practices in searching for and interpreting results of subgroup analyses. The specific focus is analyses that may show how such factors might modify intervention-outcome associations.

Some organizations, such as the Cochrane Collaboration[20] and CRD,[25] are more prescriptive than others with regard to the types of analyses that SR authors should perform to identify potential effect-measure modifiers during the analysis phase of the review. For example, Cochrane and CRD suggest meta-regression, use of I^2, and evaluation of forest plots. Perhaps because of the specific nature of the HuGENet SRs, the authors of this manual do provide suggestions for the analysis including estimating the among-study variance (I^2 statistic) and meta-regression with sensitivity analyses.

Most other organizations, including AHRQ, leave it up to the review authors to determine which analyses to conduct. A major remaining question, then, is exactly how dogmatic or laissez faire any review group should be with respect to dictating how analyses should be conducted. The desirable balance lies between permitting creative analyses that may uncover useful information, even if the analyses are unplanned at the outset, against protecting against analyses that may go beyond what can be defended conceptually, biologically, or statistically.

Critiques of How Systematic Reviews Handle Clinical Heterogeneity

Peer and Public Review Comments about AHRQ Draft Reports

For another perspective on how SRs deal with clinical heterogeneity and effect-measure modifiers, we reviewed the peer and public review comments for three *draft* AHRQ-supported CERs:

- *Comparative Effectiveness of Drug Therapy for Rheumatoid Arthritis and Psoriatic Arthritis in Adults*[12]
- *Comparative Effectiveness of Percutaneous Coronary Bypass Grafting for Coronary Artery Diseases*[13]
- *Comparative Effectiveness of Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression.*[52]

These types of reviews have been managed in some ways that differ from the approach for typical SRs (which did not necessarily concentrate on comparisons per se); moreover, because they were initiated through the highly visible Medicare Modernization Act that established the Effective Health Care Program at the Agency, CERs were subject to far more public scrutiny. This included, therefore, not just expert, external peer review but also public comment administered through a website managed at the Oregon Health & Science University EPC. Our examination of the peer and public comments for these three reviews was limited to the comments themselves; we did not also explore the "peer review disposition reports" that EPCs prepared in response to the comments, which would indicate what revisions to the draft they might (or might not) have made, depending on the salience and accuracy of the comments in question.

Reviewers for all three draft reports indicated that information was lacking on clinically relevant subgroups. Other critiques focused on the completeness and appropriateness of the studies included in the review, handling of subgroup analyses, and interpretation and presentation of clinical heterogeneity analysis. Reviewers offered these comments despite the fact that all three draft reports did have specific questions about factors that could lead to clinical heterogeneity.

Several explanations for the comments are possible. First, studies to address the specific peer or public review concern may simply have been unavailable (i.e., never conducted; never published). Second, some studies may have been excluded from the final analyses because they had not met initial inclusion criteria or had been rated "poor" for internal validity (risk of bias). Third, reviewers may have misunderstood the initial KQs that guided the review or the ultimate review. Fourth, reviewers may have been correct in their assessments of deficiencies in the draft reviews. These critiques could be addressed through more explicit discussion of what literature is and is not available for consideration of clinical heterogeneity issues. This could occur in both the results and discussion sections.

Our assessment of reasons for these criticisms led us to conclude that authors of SRs must provide more information to allay concern about incomplete or uninformative data, even when information on important clinical factors is not available. In addition, when such information is available but not included, authors need to be explicit in describing their techniques and decisions for why the information was not included in the report.

Critiques in the Literature

In addition to the peer and public critiques of draft AHRQ reports just discussed, we also identified themes in the literature about how SR authors address, or fail to address, clinical heterogeneity. Of the 83 papers that we summarized in Evidence Table C3, 24[4,5,11,38,166,167-185] noted that data dredging might result in spurious findings, which is especially salient when the analysis of the original studies disregards the randomized nature of the study design. Authors tended to disparage post hoc identification of subgroups even though some of the key informants we interviewed contended that the literature for many clinical questions is so sparse that a priori identification is almost impossible.

Similarly, our literature scan noted that the evaluation of many possible effect-measure modifiers by primary authors affects the overall type I error probability, meaning that a null hypothesis of no difference or no change is falsely rejected and that a spurious "positive" finding is reported. This may be a particular risk when evaluating many factors or subgroups. Given the aim of providing information for later clinical (or policy) decisions, this is an error systematic reviewers would like to avoid. Thus, limiting the numbers of different subgroups analyzed will help minimize this concern.

Many of the publications did indicate that analysis of individual patient-level data in meta-analyses does allow better assessment of clinical heterogeneity. Obtaining access to these data is often difficult and time-consuming and many authors refuse to relinquish their data for these types of analyses.

Evidence for How Best To Address Clinical Heterogeneity in a Systematic Review

The literature we reviewed did not provide best practices lessons for this project, with the possible exception of emphasizing the importance of specifying potential effect-measure modifiers and subgroups a priori. (We are not here addressing the many methodologic or statistical issues about conducting SRs in general or about carrying out meta-analyses, about which considerable published guidance is available.)

One interesting problem did arise, however, that suggests some best practices rest simply in keeping concepts clear. As noted earlier with respect to definitions, clinical heterogeneity is closely linked to statistical heterogeneity.[6] One can observe statistical heterogeneity by calculating the Q statistic or I^2 or by conducting a meta-regression. Statistical heterogeneity, when it occurs, can be a product of either clinical or methodological heterogeneity, where methodologic heterogeneity refers to study quality, publication bias, and other issues regarding the study design or conduct of the individual studies included in any SR.

In examining the 83 publications referred to above, we quickly saw that some authors combined clinical and methodological heterogeneity under the rubric of clinical heterogeneity. This confusion may have delayed the recognition of how important clinical heterogeneity is to the interpretation of SR findings. In reviewing the publications that did and did not differentiate between clinical and methodological heterogeneity, we concluded that the distinction did not appear to be related to the timing of publication.

Our interviews with systematic reviewers did suggest that considering clinical heterogeneity at all stages of the review is important: as the KQs are developed, when considering the inclusion and exclusion criteria, in designing the abstraction form, in abstracting the information, and in the analysis and drafting of the final report.

Limitations of Our Review

The focus of this methods report was to understand how AHRQ and other review groups handle clinical heterogeneity in the design of any SR or CER. Attention was directed to developing KQs and to analyzing the data provided by included studies. This review thus required us to evaluate a broad base of information. This body of work included guidance manuals from a variety of SR organizations, SRs conducted by these review organizations, and a literature scan for publications that discuss issues related to clinical heterogeneity.

Not surprisingly, the literature we reviewed cannot be considered comprehensive. We did attempt to examine materials from major US and international organizations that produce authoritative SRs, but we did not attempt to retrieve guidance manuals or reviews from all possible organizations around the world that might develop such reviews. For instance, we did not seek materials from the numerous governmental technology assessment, quality improvement, or practice guidelines agencies that exist globally. Apart from the time and resource constraints for the project, we also reasoned that some (if not many) of these entities may rely on SRs done by the groups that we did target. This would mean that trying to get guidance documents or reviews from such agencies or organizations would not return information commensurate with the effort.

We did review the available guidance documents (handbooks, manuals) from the major organizations. Even within that sphere, however, we were unable to collect or analyze all the reviews that they have produced over the years. To have gone much beyond the samples of reviews we did examine might have rendered the return on investment of time and resources fairly minimal.

Given the information base we assembled, we tried to maximize the comparability of the SRs that we abstracted by looking broadly across health conditions or diseases and sources of reviews. Of course, the reviews and other materials that we examined may not reflect current opinions and procedures followed by the majority of the reviewers associated with a particular review group. Thus, we also tried to obtain current views on these topics from a small number of international leaders in the evidence-based practice arena. Generally, we did not hear from them views that were markedly divergent from what we gleaned from the examination of written materials.

Good research requires studying many people to derive robust estimates. To understand how the results apply to specific patients or types of patients, however, researchers often evaluate patient subgroups. The more subgroups included in an investigation of treatment effects, the better we are able to estimate the effect of treatment on individuals who may (or may not) fit into those subgroups.

Nevertheless, subgroup analyses will never be able to predict perfectly treatment outcomes at the individual level. Although many of the patients in a given study will share many similar or even identical characteristics, they will still vary in other ways (both measurable and immeasurable). Furthermore, even if we possessed perfect knowledge of the environmental, genetic, and other clinical factors operating on or within an individual, perfect prediction of treatment effects at the individual level would also require a perfect knowledge of the specific disease process *in that individual* (e.g., sufficient and necessary factors involved in disease progression, remission, relapse, or other events). For these reasons, although including well-conducted and disease-specific subgroup analyses as part of comparative effectiveness research should get us closer to treating at the individual level, there will be a point when additional

studies will not provide any new information to help clinicians treat or researchers try to predict outcomes for specific patients using information derived from subgroups.

For these reasons, we acknowledge that many questions about how to specify and handle clinical heterogeneity in SRs remain. The next section picks up on these points in more detail.

AHRQ EPC Work Group

The sixth KQ posed for this project called for us to lay out a set of issues that an AHRQ Work Group for the EPC program might address. Numerous methodological issues have been the focus of such work groups, particularly in terms of CERs that are done through AHRQ's Effective Health Care Program; several of these are now part of the AHRQ *EPC Methods Guide* cited earlier. Topics raised throughout our study and highlighted in the results and discussion presented to this point cut across major challenges for doing SRs in ways that meet AHRQ and, particularly, stakeholder needs. Of special concern is how clinical heterogeneity affects analyses (particularly quantitative synthesis) and applicability—broadly considered as translation of results into useful information for clinicians and policymakers.

Our review and analysis raise the following issues that we believe an EPC Work Group could usefully address further. We note, as well, that these are not settled matters in the broader world of systematic reviewers. Thus, any elucidation of these types of questions should prove of benefit beyond the AHRQ ambit. For that reason, and to gain the most up-to-date thinking across many groups dealing with these same problems, we suggest that AHRQ may wish to involve leaders in the SR field from outside AHRQ and outside the United States.

Topics for a Specific Charge to the Work Group

Table 19 provides 11 questions, offered in a somewhat "chronological order" as authors might move through an SR, as our priority recommendations for what an EPC work group might address. Many have obvious (or not so obvious) subquestions, but we believe that this set would establish a robust agenda for any work group.

Table 19. Topics for specific charge to the work group

1.	Is the definition of clinical heterogeneity clear enough for future work? Are the distinctions between it and statistical heterogeneity clear as well? Should clinical heterogeneity be distinguished from clinical diversity?
2.	Are the basic categories of clinical heterogeneity introduced for this study—demographic characteristics; clinical variables involving disease severity, stage, or type, risk factors, coexisting conditions, and cointerventions—satisfactory? Are they sufficient?
3.	What process might be developed for determining which clinical heterogeneity factors a review should consider? Would this process differ depending on whether the work (for AHRQ) is a standard systematic review or a comparative effectiveness review?
4.	Should restriction be part of the toolkit for addressing clinical heterogeneity?
5.	Should the process (question 3) mandate "only" a priori statements of clinical heterogeneity factors to be taken in account?
6.	How would such a process take account of what sponsors or nominators of topics have suggested in this context? How would it take elements of the clinical problems, health interventions, and other aspects that differ markedly across the range of reviews that AHRQ sponsors?
7.	Would such a process permit post hoc identification of subgroups for further analysis? If so, what conditions might it set for authors to justify such decisions? What role might individual studies rated "poor quality" (and likely excluded from final analyses) play when no other acceptable evidence on important subgroups exists?
8.	Do appropriate statistical tests exist for assessing clinical heterogeneity? If so, how can such information best be provided as guidance to EPC reviewers?
9.	Should a plan for clinical heterogeneity assessment be part of the posted workplan?
10.	Should each EPC systematic review include a description for how they will handle clinical heterogeneity in the methods section of the review? If so, where should a description of the findings from the clinical heterogeneity assessment be placed?
11.	What recommendations might be made for agreed-upon terminology and standard reporting of clinical heterogeneity results?

Summary

- Evaluating effect-measure modification of an intervention-outcome association is very important to a variety of audiences (i.e., practicing clinicians, patients, and policymakers, and to persons doing SRs or CERs). Planning for it ideally should be done a priori in the protocol development phase, although identifying subgroups for analysis post hoc remains an open question.
- We found nearly identical definitions of clinical heterogeneity from various organizations framing the idea as variability in populations, interventions and outcomes. In-depth review suggests that conceptualization of clinical heterogeneity differs among researchers, however, with some considering clinical heterogeneity to include methodologic heterogeneity. We and others do not include methodologic heterogeneity within the rubric of clinical heterogeneity. Future work to develop consensus around common definitions of heterogeneity (clinical vs. statistical) and to explore restriction more broadly as a way to handle clinical heterogeneity would be highly valuable.
- In the protocol development and analysis phases of a review, any assessment of clinical heterogeneity frequently includes demographic factors, coexisting conditions and/or cointerventions; disease stage or severity, risk factors, or all these factors are included less often. Inclusion of less common variables or more narrowly defined subgroups in reviews likely reflects limited availability in the included literature.
- In order for an SR or CER to evaluate which factors influence an intervention-treatment effect requires sufficient data addressing these factors in the individual studies in the SRs and CERs. However, researchers often do not pre-specify which subgroups to evaluate when designing their studies, which preclude powering the study adequately to

investigate potential effect-measure modification. Further, even if subgroups have been evaluated, the investigators may not report these analyses because they are too small to provide robust conclusions and may be seen as data dredging.

- Most groups recommend including a comprehensive description of the process used in identifying and selecting factors for assessment of clinical heterogeneity, with several suggesting the use of specific tools and techniques (forest plots, pooling, meta-regression, etc.), where applicable.
- With only one exception, groups did not document their use of experts and/or literature as guiding development of KQs. We would recommend the use of clinical experts and/or literature to identify factors related to clinical heterogeneity, and doing so a priori, as supported by existing guidance.
- How flexible reviewers should be in conducting subgroup analyses remains unclear. The balance lies somewhere between permitting creative analyses, even if unplanned, to uncover useful information, yet avoiding analyses that cannot be defended conceptually, biologically, or statistically.

References

1. Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. J Am Med Assoc 2010;303:985-986.
2. Rothwell PM. Can overall results of clinical trials be applied to all patients? Lancet 1995 Jun 24;345(8965):1616-1619.
3. Santaguida PL, Helfand M, Raina P. Challenges in systematic reviews that evaluate drug efficacy or effectiveness. Annals of Internal Medicine 2005 Jun 21;142(12 Pt 2):1066-1072.
4. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science Epidemiologic Reviews 1992;14:154-176.
5. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. Am J Epidemiol 1995 Aug;142(4):383-387.
6. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994 Nov 19;309(6965):1351-1355.
7. West SL, Garbutt JC, Carey TS, et al. Pharmacotherapy for alcohol dependence. Evidence report number 3. (Contract 290-97-0011 to Research Triangle Institute, University of North Carolina, Chapel Hill). AHCPR publication no. 99-E004. Rockville, MD: Agency for Health Care Policy and Research 1999.
8. Garbutt JC, West SL, Carey TS, et al. Pharmacological treatment of alcohol dependence: a review of the evidence. J Am Med Assoc 1999;281:1318-1325.
9. Gartlehner G, Thieda P, Hansen RA, et al. Comparative risk for harms of second-generation antidepressants: a systematic review and meta-analysis. Drug Saf 2008;31(10):851-865.
10. Morton SC, Adam JL, Suttorp MJ, et al. Meta-regression Approaches: What, Why, When, and How? Technical Review 8 prepared by Southern California-RAND-Evidence-based Practice Center, under Contract No 290-97-0001. AHRQ Publication No 04-033, Rockville, MD: Agency for Healthcare Research and Quality 2004.
11. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. Am J Epidemiol 1995 Aug 15;142(4):371-382.
12. Donahue KE, Gartlehner G, Jonas DE, et al. Systematic review: comparative effectiveness and harms of disease-modifying medications for rheumatoid arthritis. Ann Intern Med 2008 Jan 15;148(2):124-134.
13. Bravata DM, Gienger AL, McDonald KM, et al. Systematic review: the comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery. Ann Intern Med 2007 Nov 20;147(10):703-716.
14. Brouwer RW, van Raaij TM, Bierma-Zeinstra SMA, et al. Osteotomy for treating knee osteoarthritis. Cochrane Database of Systematic Reviews: Reviews 2007;Issue 3.
15. Christensen R, Bartels EM, Astrup A, et al. Symptomatic efficacy of avocado-soybean unsaponifiables (ASU) in osteoarthritis (OA) patients: a meta-analysis of randomized controlled trials. Osteoarthritis and Cartilage 2008 Apr;16(4):399-408.
16. Li J, Zhang Q, Zhang M, et al. Intravenous magnesium for acute myocardial infarction. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.
17. Sinno MC, Khanal S, Al-Mallah MH, et al. The efficacy and safety of combination glycoprotein IIbIIIa inhibitors and reduced-dose thrombolytic therapy-facilitated percutaneous coronary intervention for ST-elevation myocardial infarction: a meta-analysis of randomized clinical trials. American Heart Journal 2007;153(4):579-586.
18. Magee K, Campbell Samuel G, Moher D, et al. Heparin versus placebo for acute coronary syndromes. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.
19. Samson DJ, Grant MD, Ratko TA, et al. Treatment of primary and secondary osteoarthritis of the knee. Rockville: Agency for Healthcare Research and Quality (AHRQ) 2007.

20. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1. Available at http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf. Accessed 2008.
21. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005 Jan 8-14;365(9454):176-186.
22. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Available at http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Accessed 2007.
23. Project DER. Review methods and report production procedures. Available at http://www.ohsu.edu/ohsuedu/researchpolicycenter/DERP/. Accessed May 28, 2009.
24. The HuGENet Writing Group. The HuGENET™ HuGE Review Handbook, Version 1.0. 06. Available at http://www.hugenet.ca. Accessed February 28, 2009.
25. Centre for Reviews and Dissemination. Centre for Reviews and Disssemination's (CRD) Systematic Reviews: Guidance for undertaking reviews in health care. Available at http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf. Accessed March 9, 2009.
26. National Institute for Health and Clinical Excellence. The guidelines manual. Available at http://www.nice.org.uk/media/5F2/44_The_guidelines_manual_2009_-_All_chapters.pdf. Accessed October 19, 2009.
27. Institute for Quality and Efficiency in Health Care. General Methods. Available at http://www.iqwig.de/download/IQWiG_General_methods_V-3-0.pdf. Accessed October 19, 2009.
28. National Health and Medical Research Council. How to review the evidence: systematic identification and review of the scientific literature. Available at http://www.nhmrc.gov.au/_files_nhmrc/file_publications/synopses/cp65.pdf. Accessed October 9, 2009.
29. European Network for Health Technology Assessment. HTA core model for medical and surgical interventions 1.0R. Available at http://www.eunethta.net/upload/WP4/Final%20Deliverables/HTA%20Core%20Model%20for%20Medical%20and%20Surgical%20Interventions%201%200r.pdf. Accessed March 9, 2009.
30. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1, Chapter 9 (page 9.27). Available at http:www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf. Accessed November 13, 2009.
31. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002 Jun 12;287(22):2973-2982.
32. Schulz KF. Meta-analyses of interventional trials done in populations with different risks. Lancet 1995 May 20;345(8960):1304-1305.
33. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Health Care Program. J Clin Epidemiol 2009 Jul 10.
34. Cochran W. The combination of estimates from different experiments Biometrics 1954;10:101-129.
35. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558.
36. Greenland S, O'Rourke K. Meta-Analysis. In: Rothman KJ, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins 2008:652-682.
37. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560.
38. Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002 Feb 15;21(3):371-387.

39. Engels EA, Schmid CH, Terrin N, et al. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med 2000 Jul 15;19(13):1707-1728.

40. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. American Journal of Epidemiology 1999 Sep;150(5):469-475.

41. Shadish WR, Cook, TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin College Div 2003.

42. Citrome L. Compelling or irrelevant? Using number needed to treat can help decide. Acta Psychiatr Scand 2008 Jun;117(6):412-419.

43. Steg PG, Lopez-Sendon J, Lopez de Sa E, et al. External validity of clinical trials in acute myocardial infarction. Arch Intern Med 2007 Jan 8;167(1):68-73.

44. Zarin DA, Young JL, West JC. Challenges to evidence-based medicine: a comparison of patients and treatments in randomized controlled trials with patients and treatments in a practice research network. Soc Psychiatry Psychiatr Epidemiol 2005 Jan;40(1):27-35.

45. Dhruva SS, Redberg RF. Variations between clinical trial participants and Medicare beneficiaries in evidence used for Medicare national coverage decisions. Arch Intern Med 2008 Jan 28;168(2):136-140.

46. Bruening W, Launders J, Pinkney N, et al. Effectiveness of noninvasive diagnostic tests for breast abnormalities. Comparative Effectiveness Review No. 2. (Prepared by ECRI Evidence-based Practice Center under Contract No. 290-02-0019.) Rockville, MD: Agency for Healthcare Research and Quality. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm 2006.

47. Seidenfeld J, Samson DJ, Bonnell CJ, et al. Management of small cell lung cancer. Evidence Report/Technology Assessment No. 143. (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-02-0026.) AHRQ Publication No. 06-E016. Rockville, MD: Agency for Healthcare Research and Quality 2006.

48. Wilt TJ, Shamliyan T, Taylor B, et al. Comparative effectiveness of therapies for clinically localized prostate cancer. Comparative Effectiveness Review No. 13. (Prepared by Minnesota Evidence-based Practice Center under Contract No. 290-02-0009.) Rockville, MD: Agency for Healthcare Research and Quality. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm 2008.

49. Shekelle P, Rich M, Morton S, et al. Pharmacologic management of heart failure and left ventricular systolic dysfunction: effect in female, black, and diabetic patients, and cost-effectiveness. Evidence Report/Technology Assessment No. 82 (Prepared by the Southern California-RAND Evidence-based Practice Center under Contract No 290-97-0001). AHRQ Publication No. 03-E045. Rockville, MD: Agency for Healthcare Research and Quality. July 2003.

50. Viswanathan M, Visco AG, Hartmann K, et al. Cesarean delivery on maternal request. Evidence Report/Technology Assessment No. 133. AHRQ Pub. No. 06-E009. Rockville, MD: Agency for Healthcare Research and Quality. 2006.

51. Wilt TJ ND, Kim C-B, Kane RL, et al. Use of spirometry for case finding, diagnosis, and management of chronic obstructive pulmonary disease (COPD). Evidence Report/Technology Assessment No. 121 (Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-02-0009.) AHRQ Publication No. 05-E017-2. Rockville, MD. Agency for Healthcare Research and Quality. 2005.

52. Gartlehner G, Hansen RA, Thieda P, et al. Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression. Comparative Effectiveness Review No. 7. (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.) Rockville, MD: Agency for Healthcare Research and Quality. January 2007. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm. 2007.

53. Ip S, Bonis P, Tatsioni A, et al. Comparative effectiveness of management strategies for gastroesophageal reflux disease. Comparative Effectiveness Review No. 1. (Prepared by Tufts-New England Medical Center Evidence-based Practice Center under Contract No. 290-02-0022.) Rockville, MD: Agency for Healthcare Research and Quality. December 2005. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm. 2005.

54. Matchar DB, McCrory DC, Orlando LA, et al. Comparative effectiveness of angiotensin-converting enzyme Inhibitors (ACEIs) and angiotensin II receptor antagonists (ARBs) for treating essential hypertension. Comparative Effectiveness Review No. 10. (Prepared by Duke Evidence-based Practice Center under Contract No. 290-02-0025.) Rockville, MD: Agency for Healthcare Research and Quality. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm. 2007.

55. Caughey AB, Sundaram V, Kaimal AJ, et al. Maternal and neonatal outcomes of elective induction of labor. Evidence Report/Technology Assessment No. 176. (Prepared by the Stanford University-UCSF Evidenced-based Practice Center under contract No. 290-02-0017.) AHRQ Publication No. 09-E005. Rockville, MD: Agency for Healthcare Research and Quality. 2009.

56. Chou R, Helfand M, Peterson K, et al. Comparative effectiveness and safety of analgesics for osteoarthritis. Comparative Effectiveness Review No. 4. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-02-0024.) Rockville, MD: Agency for Healthcare Research and Quality. September 2006. Available at http://www.effectivehealthcare.ahrq.gov/reports/final.cfm 2006.

57. Zhang M, Liu X, Li J, et al. Chinese medicinal herbs to treat the side-effects of chemotherapy in breast cancer patients. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

58. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365(9472):1687-1717.

59. Amarasena IU, Walters JAE, Wood-Baker R, et al. Platinum versus non-platinum chemotherapy regimens for small cell lung cancer. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 4.

60. Cardona Zorrilla AF, Reveiz L, Ospina Edgar G, et al. Palliative endobronchial brachytherapy for non-small cell lung cancer. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 2.

61. Dong R, Chen X, Wu T, et al. Elemene for the treatment of lung cancer. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

62. Shelley M, Wilt T, Coles B, et al. Cryotherapy for localised prostate cancer. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

63. De Conti P, Atallah Álvaro N, Arruda Homero O, et al. Intermittent versus continuous androgen suppression for prostatic cancer. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

64. Frobel A-K, Hulpke-Wette M, Schmidt Klaus G, et al. Beta-blockers for congestive heart failure in children. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2009.

65. Mathai M, Hofmeyr GJ. Abdominal surgical incisions for caesarean section. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

66. Anorlu RI, Maholwana B, Hofmeyr GJ. Methods of delivering the placenta at caesarean section. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

67. Navaneethan SD, Pansini F, Perkovic V, et al. HMG CoA reductase inhibitors (statins) for people with chronic kidney disease not requiring dialysis. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2009.

68. Roderick PJ, Willis NS, Blakeley S, et al. Correction of chronic metabolic acidosis for chronic kidney disease patients. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

69. Effing T, Monninkhof EM, van der Valk P, et al. Self-management education for patients with chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews: Reviews. 2007;Issue 4.

70. Puhan M, Scharplatz M, Troosters T, et al. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews: Reviews. 2009;Issue 1.

71. Puhan MA, Vollenweider D, Latshang T, et al. Exacerbations of chronic obstructive pulmonary disease: when are antibiotics indicated? A systematic review. Respiratory Research 2007;8:30.

72. Yang IA, Fong K, Sim EHA, et al. Inhaled corticosteroids for stable chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

73. Nieuwenhuijsen K, Bültmann U, Neumeyer-Gromen A, et al. Interventions to improve occupational health in depressed people. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

74. Morriss R, Faizal Mohammad A, Jones Ashley P, et al. Interventions for helping people recognise early signs of recurrence in bipolar disorder. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

75. Hackett ML, Anderson CS, House A, et al. Interventions for preventing depression after stroke. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

76. Furtado Vivek A, Srihari V, Kumar A. Atypical antipsychotics for people with both schizophrenia and depression. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

77. Cipriani A, Rendell Jennifer M, Geddes J. Olanzapine in long-term treatment for bipolar disorder. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2009.

78. Liu Jian P, Yang H, Xia Y, et al. Herbal preparations for uterine fibroids. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2009.

79. Beaumont HH, Augood C, Duckitt K, et al. Danazol for heavy menstrual bleeding. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

80. Lethaby A, Augood C, Duckitt K, et al. Nonsteroidal anti-inflammatory drugs for heavy menstrual bleeding. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

81. Lethaby A, Irvine Gill A, Cameron Iain T. Cyclical progestogens for heavy menstrual bleeding. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

82. Wiysonge C, Shey U, Bradley HA, et al. Beta-blockers for hypertension. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

83. Woolf SH, Battista RN, Anderson GM, et al. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. A report by the Canadian Task Force on the Periodic Health Examination. J Clin Epidemiol 1990;43(9):891-905.

84. Abalos E, Duley L, Steyn DW, et al. Antihypertensive drug therapy for mild to moderate hypertension during pregnancy. Cochrane Database of Systematic Reviews 2007(1):CD002252.

85. Zijdenbos IL, de Wit Niek J, van der Heijden GJ, et al. Psychological treatments for the management of irritable bowel syndrome. Cochrane Database of Systematic Reviews: Reviews. 2009;Issue 1.

86. Webb AN, Kukuruzovic R, Catto-Smith AG, et al. Hypnotherapy for treatment of irritable bowel syndrome. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

87. Evans Beti W, Clark Wendy K, Moore David J, et al. Tegaserod for the treatment of irritable bowel syndrome and chronic constipation. Cochrane Database of Systematic Reviews: Reviews. 2007;Issue 4.

88. Smyth R, Alldred SK, Markham C. Amniotomy for shortening spontaneous labour. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2007.

89. Boulvain M, Kelly Anthony J, Irion O. Intracervical prostaglandins for induction of labour. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 1.

90. Wu T, Ni J, Wei J. Danshen (Chinese medicinal herb) preparations for acute myocardial infarction. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

91. Fransen M, McConnell S. Exercise for osteoarthritis of the knee. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

92. Laupattarakasem W, Laopaiboon M, Laupattarakasem P, et al. Arthroscopic debridement for knee osteoarthritis. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. 2008.

93. Boutin DL. Effectiveness of cognitive behavioral and supportive-expressive group therapy for women diagnosed with breast cancer: a review of the literature. Journal for Specialists in Group Work. 2007;32(3):267-284.

94. van der Ploeg IM, van Esser S, van den Bosch MA, et al. Radiofrequency ablation for breast cancer: a review of the literature. European Journal of Surgical Oncology. 2007;33(6):673-677.

95. Yau G, Lock M, Rodrigues G. Systematic review of baseline low-dose CT lung cancer screening. Lung Cancer. 2007;58(2):161-170.

96. Micames CG, McCrory DC, Pavey DA, et al. Endoscopic ultrasound-guided fine-needle aspiration for non-small cell lung cancer staging: a systematic review and metaanalysis. Chest. 2007;131(2):539-548.

97. Coory M, Gkolia P, Yang IA, et al. Systematic review of multidisciplinary teams in the management of lung cancer. Lung Cancer. 2008;60(1):14-21.

98. Tarnhuvud M, Wandel C, Willman A. Nursing interventions to improve the health of men with prostate cancer undergoing radiotherapy: a review. European Journal of Oncology Nursing. 2007;11(4):328-339.

99. Lago RM, Singh PP, Nesto RW. Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials. Lancet 2007;370:1129-1136.

100. Nasr IA, Bouzamondo A, Hulot JS, et al. Prevention of atrial fibrillation onset by beta-blocker treatment in heart failure: a meta-analysis. European Heart Journal 2007;28(4):457-462.

101. Roberts F, Ryan GJ. The safety of metformin in heart failure. Annals of Pharmacotherapy. 2007;41(4):642-646.

102. Nicholson WK, Wilson LM, Witkop CT, et al. Therapeutic management, delivery, and postpartum risk assessment and screening in gestational diabetes. Report 2008:96.

103. Press JZ, Klein MC, Kaczorowski J, et al. Does cesarean section reduce postpartum urinary incontinence: a systematic review. Birth 2007;34(3):228-237.

104. Balamuthusamy S, Srinivasan L, Verma M, et al. Renin angiotensin system blockade and cardiovascular outcomes in patients with chronic kidney disease and proteinuria: a meta-analysis. American Heart Journal. 2008;155(5):791-805.

105. Strippoli GF, Navaneethan SD, Johnson DW, et al. Effects of statins in patients with chronic kidney disease: meta-analysis and meta-regression of randomised controlled trials. BMJ 2008;336:645-651.

106. Bradley-Drummond M, Dasenbrook EC, Pitz MW, et al. Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: a systematic review and meta-analysis. JAMA 2008;300(20):2407-2016.

107. Niesink A, Trappenburg JC, de Weert-van Oene GH, et al. Systematic review of the effects of chronic disease management on quality-of-life in people with chronic obstructive pulmonary disease. Respiratory Medicine 2007;101(11):2233-2239.

108. Quon BS, Gan WQ, Sin DD. Contemporary management of acute exacerbations of COPD: a systematic review and metaanalysis. Chest 2008;133(3):756-766.

109. Singh S, Loke YK, Furberg CD. Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. JAMA 2008;300(12):1439-1450.

110. Sobieraj DM, White CM, Coleman CI. Benefits and risks of adjunctive inhaled corticosteroids in chronic obstructive pulmonary disease: a meta-analysis. Clinical Therapeutics. 2008;30(8):1416-1425.

111. Usala T, Clavenna A, Zuddas A, et al. Randomised controlled trials of selective serotonin reuptake inhibitors in treating depression in children and adolescents: a systematic review and meta-analysis. European Neuropsychopharmacology 2008;18(1):62-73.

112. Cuijpers P, Brannmark JG, van Straten A. Psychological treatment of postpartum depression: a meta-analysis. Journal of Clinical Psychology 2008;64(1):103-118.

113. Chin AM. Clinical effects of reminiscence therapy in older adults: a meta-analysis of controlled trials. Hong Kong Journal of Occupational Therapy 2007;17(1):10-22.

114. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. Canadian Medical Association Journal 2008;178(3):296-305.

115. Hiyama T, Yoshihara M, Matsuo K, et al. Meta-analysis of the effects of prokinetic agents in patients with functional dyspepsia. Journal of Gastroenterology and Hepatology 2007;22(3):304-310.

116. Jin X, Li YM. Systematic review and meta-analysis from Chinese literature: the association between Helicobacter pylori eradication and improvement of functional dyspepsia. Helicobacter 2007;12(5):541-546.

117. Wang WH, Huang JQ, Zheng GF, et al. Effects of proton-pump inhibitors on functional dyspepsia: a meta-analysis of randomized placebo-controlled trials. Clinical Gastroenterology and Hepatology 2007;5(2):178-185.

118. Horvath K, Jeitler K, Siering U, et al. Long-term effects of weight-reducing interventions in hypertensive patients: systematic review and meta-analysis. Archives of Internal Medicine 2008;168(6):571-580.

119. Connell P, Wolfe C, McKevitt C. Preventing stroke: a narrative review of community interventions for improving hypertension control in black adults. Health and Social Care in the Community 2008;16(2):165-187.

120. Ford AC, Talley NJ, Velhuyzen van Zenten SJ, et al. Will the history and physical examination help establish that irritable bowel syndrome is causing this patient's lower gastrointestinal tract symptoms? JAMA. 2008;300(15):1793-1805.

121. Ford AC, Talley NJ, Spiegel BM, et al. Effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome: systematic review and meta-analysis. BMJ 2008;337:a2313.

122. Ford AC, Chey WD, Talley NJ, et al. Yield of diagnostic tests for celiac disease in individuals with symptoms suggestive of irritable bowel syndrome. Archives of Internal Medicine 2009;169(7):651-658.

123. Rahimi R, Nikfar S, Abdollahi M. Selective serotonin reuptake inhibitors for the management of irritable bowel syndrome: a meta-analysis of randomized controlled trials. Archives of Medical Science 2008;4(1):71-76.

124. Baker WL, Couch KA. Azithromycin for the secondary prevention of coronary artery disease: a meta-analysis. American Journal of Health-System Pharmacy 2007;64(8):830-836.

125. Ioannidis JP, Katritsis DG. Percutaneous coronary intervention for late reperfusion after myocardial infarction in stable patients. American Heart Journal 2007;154(6):1065-1071.

126. Minns Lowe CJ, Barker KL, Dewey M, et al. Effectiveness of physiotherapy exercise after knee arthroplasty for osteoarthritis: systematic review and meta-analysis of randomised controlled trials. BMJ 2007;335:812.
127. Pisters MF, Veenhof C, van Meeteren NL, et al. Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review. Arthritis and Rheumatism 2007;57(7):1245-1253.
128. Chou R, Helfand M, Carson S. Drug class review on angiotensin converting enzyme inhibitors. Final report. Portland, OR: Oregon Health & Science University. 2005.
129. Furmaga E, Glassman P, Rhodes S, et al. Drug class review on angiotensin II receptor antagonists. Final report. Portland, OR: Oregon Health & Science University. 2006.
130. Helfand M, Peterson K, Christensen V, et al. Drug class review on beta adrenergic blockers. Final report update 4. Portland, OR: Oregon Health & Science University. 2009.
131. Jonas DE, Kiser K, Shilliday BB, et al. Drug class review on controller medications for asthma. Portland, OR: Oregon Health & Science University. 2008.
132. Hansen RA, Gartlehner G, Lohr KN, et al. Drug class review on inhaled corticosteroids. Final report update 1. Portland, OR: Oregon Health & Science University. 2006.
133. Gartlehner G, Morgan LC, Thieda P, et al. Drug class review second generation antidepressants. Final report update 4. Portland, OR: Oregon Health & Science University. 2008.
134. McDonagh MS, Carson S, Thakurta S. Drug class review on proton pump inhibitors. Final report update 5. Portland, OR: Oregon Health & Science University. 2009.
135. McDonagh MS, Eden KB, Peterson K. Drug class review on calcium channel blockers. Final report. Portland, OR: Oregon Health & Science University. 2005.
136. Gartlehner G, Jonas DE, Morgan LC, et al. Drug class review on constipation drugs. Final report. Portland, OR: Oregon Health & Science University. 2007.
137. Dailey JH, Glassman P, McDonagh M, et al. Drug class review on newer antiplatelet agents. Final report update 1. Portland, OR: Oregon Health & Science University. 2007.
138. Helfand M, Carson S, Kelley C. Drug class review on HMG-CoA reductase inhibitors (statins). Final report. Portland, OR: Oregon Health & Science University. 2006.
139. Chou R, Helfand M, Peterson K, et al. Drug class review on cyclo-oxygenase (COX)-2 inhibitors and non-steroidal anti-inflammatory drugs (NSAIDs). Final report update 3. Portland, OR: Oregon Health & Science University. 2006.
140. Adelaide Health Technology A. Digital mammography for breast cancer screening and diagnosis. Adelaide: Adelaide Health Technology Assessment (AHTA). 2008.
141. Korencan A, Felder-Puig R, Guba B, et al. Testing for HER2 positive breast cancer. Challenge for improvement of current conditions and practice. Wien: Ludwig Boltzmann Institut fuer Health Technology Assessment (LBIHTA). 2007.
142. Dunfield L, Severn M. Effectiveness of magnetic resonance imaging (MRI) screening for women at high risk of breast cancer. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH). 2007.
143. Adelaide Health Technology Assessment on behalf of National Horizon Scanning U. Genetic Screening for Lung Cancer utilising CT; horizon scanning prioritizing summary - volume 15. Adelaide: Adelaide Health Technology Assessment (AHTA) on behalf of National Horizon Scanning Unit (HealthPACT and MSAC). 2007.
144. Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen. Interstitial brachytherapy in localised prostate cancer. Cologne: Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). 2007.
145. Pearson SD, Ladapo J, Prosser L. Intensity modulated radiation therapy (IMRT) for localized prostate cancer. Boston: Institute for Clinical and Economic Review (ICER). 2007.

146. Swedish Council on Technology Assessment in Health C. Light therapy for depression, and other treatment of seasonal affective disorder. Stockholm: Swedish Council on Technology Assessment in Health Care (SBU). 2007.

147. Swedish Council on Technology Assessment in Health C. Computer-based cognitive behavioral therapy for anxiety disorders or depression - early assessment briefs (Alert). Stockholm: Swedish Council on Technology Assessment in Health Care (SBU). 2007.

148. McLoughlin DM, Mogg A, Eranti S, et al. The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis. Health Technol Assess 2007 Jul;11(24):1-54.

149. Screening for postnatal depression within the Well Child Tamariki Ora Framework. Christchurch: Health Services Assessment Collaboration (HSAC). 2008.

150. National Collaborating Centre for Mental Health. Depression: management of depression in primary and secondary care. Rushden, Northamptsonshire: The British Psychological Society and Gaskell. 2004.

151. North of England Dyspepsia Guideline Development Group. Dyspepsia: managing dyspepsia in adults in primary care. Newcastle upon Tyne (UK): Centre for Health Services Research School of Population and Health Sciences University of Newcastle upon Tyne, Report No. 112 2004.

152. Cooper A, Skinner J, Nherera L, et al. Clinical guidelines and evidence review for post myocardial infarction: secondary prevention in primary and secondary care for patients following a myocardial infarction. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners. 2007.

153. National Collaborating Centre for Chronic Conditions. Osteoarthritis: national clinical guideline for care and management in adults. London: Royal College of Physicians. 2008.

154. National Collaborating Centre for Cancer. Early and locally advanced breast cancer: diagnosis and treatment. NICE Clinical Guideline. London (UK): National Institute for Health and Clinical Excellence (NICE) 2009:37.

155. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989 Dec 20;81(24):1879-1886.

156. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. Am J Hum Genet 1998 Jan;62(1):145-158.

157. Swedish Council on Technology Assessment in Health C. Dyspepsia and gastro-oesophageal reflux. Stockholm: Swedish Council on Technology Assessment in Health Care (SBU). 2007.

158. Hermanowski T, Kowalik E, Jakubczyk M, et al. Use of bosentan, epoprostenol, iloprost, sildenafil and treprostinil in treatment of pulmonary arterial hypertension in Poland. Budget impact analysis. Warsaw: Agency for Health Technology Assessment in Poland (AHTAPol). 2007.

159. Hermanowski T, Kowalik E, Jakubczyk M, et al. Treatment of pulmonary arterial hypertension in Poland - current practice. Survey results. Warsaw: Agency for Health Technology Assessment in Poland (AHTAPol). 2007.

160. Tran K, Ho C, Noorani HZ, et al. Overview of thiazide diuretics as first-line treatment for hypertension. [Technology overview number 34]. Ottawa: Canadian Agency for Drugs and Technologies in Health. 2007.

161. Comparison of efficacy, effectiveness and safety of fibrinolysis and primary percutaneous coronary intervention (PPCI) for ST-segment elevation myocardial infarction (STEMI). Montreal: Agence d'Evaluation des Technologies et des Modes d'Intervention en Sante (AETMIS). 2008.

162. National Coordinating Centre for Health Technology Assessment. The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Home-based compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence. 2007.

163. National Institute for Health and Clinical Excellence. Secondary prevention in primary and secondary care for patients following a myocardial infarction. London: National Institute for Health and Clinical Excellence (NICE). 2007.

164. Canadian Agency for Drugs and Technologies in Health (CADTH). Intra-articular hyaluronic acid (viscosupplementation) for hip osteoarthritis. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH). 2007.

165. Gartlehner G, Hansen RA, Nissman D, et al. A simple and valid tool distinguished efficacy from effectiveness studies. J Clin Epidemiol. 2006;60(7):753-755.

166. Adams KFJ. Post hoc subgroup analysis and the truth of a clinical trial [editorial]. American Heart Journal. 1998 Nov;136(5):753-758.

167. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA. 1999 Mar 3;281(9):830-834.

168. Berry DA. Subgroup analyses [letter]. Biometrics. 1990 Dec;46(4):1227-1230.

169. Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. Stat Med. 1992 Dec;11(16):2077-2082.

170. Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technology Assessment. 2001;5(33):1-56.

171. Chalmers TC, Lau J. Changes in clinical trials mandated by the advent of meta-analysis. Stat Med. 1996 Jun 30;15(12):1263-1268; discussion 9-72.

172. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol. 1995 Jan;48(1):167-171.

173. Hahn S, Williamson PR, Hutton JL, et al. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med. 2000 Dec 30;19(24):3325-3336.

174. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Statistics in Medicine. 1998 Apr;17(8):841-856.

175. Koopman L, van der Heijden GJ, Glasziou PP, et al. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. J Clin Epidemiol. 2007 Oct;60(10):1002-1009.

176. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997 Nov 1;127(9):820-826.

177. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999 Nov;354(9193):1896-1900.

178. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). Annual Review of Public Health 1996;17:1-23.

179. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992 Jan 1;116(1):78-84.

180. Reade MC, Delaney A, Bailey MJ, et al. Bench-to-bedside review: avoiding pitfalls in critical care meta-analysis--funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy. Critical Care 2008;12(4):220.

181. Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. Journal of Evaluation in Clinical Practice 2005 Oct;11(5):468-478.

182. Song FJ, Sheldon TA, Sutton AJ, et al. Methods for exploring heterogeneity in meta-analysis. Evaluation & the Health Professions 2001 Jun;24(2):126-151.

183. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;Jun 15(21):1559-1573.

184. Winegardner ML, Reaume KT, Dabaja GS, et al. Practical application of nonrandomized research to patient care: A case study of nesiritide. Pharmacotherapy 2007 Jan;27(1):143-151.

185. Xu HR, Platt RW, Luo ZC, et al. Exploring heterogeneity in meta-analyses: needs, resources and challenges. Paediatr Perinat Epidemiol 2008 Jan;22:18-28.

186. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy [comment]. Journal of Clinical Epidemiology 1998 Apr;51(4):297-299.

187. Bigger JT. Issues in subgroup analyses and meta-analyses of clinical trials. J Cardiovasc Electrophysiol 2003 Sep;14(9 Suppl):S6-S8.

188. Freemantle N, Mason J, Eccles M. Deriving treatment recommendations from evidence within randomized trials - The role and limitation of meta-analysis. International Journal of Technology Assessment in Health Care 1999 Spr;15(2):304-315.

189. Ioannidis JP, Lau J. Uncontrolled pearls, controlled evidence, meta-analysis and the individual patient. J Clin Epidemiol 1998 Aug;51(8):709-711.

190. Baujat B, Mahe C, Pignon JP, et al. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652.

191. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Stat Med 2008 Feb 28;27(5):625-650.

192. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. International Journal of Epidemiology 1999 Feb;28(1):1-9.

193. Eysenck HJ. Meta-analysis and its problems. BMJ 1994 Sep 24;309(6957):789-792.

194. Schmid CH, Lau J, McIntosh MW, et al. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 1998 Sep 15;17(17):1923-1942.

195. Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials [book chapter]. In: Egger M, Smith GD, Altman DG, eds. Systematic Review in Health Care: Meta-analysis in Context. 2nd ed. London, England: BMJ Publishing Group. 2001:143-156.

196. Chalmers TC. Problems induced by meta-analyses. Stat Med 1991 Jun;10(6):971-979; discussion 9-80.

197. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. The Cochrane Library, Issue 4. Chichester, UK: John Wiley & Sons, Ltd. 2006.

198. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Onkologie 2000 Dec;23(6):597-602.

199. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097.

200. Harris RP, Helfand M, Woolf SH, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. Am J Prev Med 2001 Apr;20(3 Suppl):21-35.

201. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ) 2002 Mar(47):1-11.

202. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002 Jun 15;21(11):1559-1573.

203. Gesundheitswesen IfQuWi. General methods. Available at http://www.iqwig.de/downloadIQWiG_General_methods_V-3-0.pdf. Accessed October 19, 2009.

Appendix A. Search Strategies

Search strategy for guidance on or best practices for addressing clinical heterogeneity

Search	Queries	Result
#37	Search #21 AND #36	1,065
#36	Search #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35	361,519
#35	Search lancet[TA]	11,9063
#34	Search PLoS ONE[ta]	5,389
#33	Search PLoS Med[ta]	1,379
#32	Search statistics in medicine[ta]	5,329
#31	Search preventive medicine[ta]	4,135
#30	Search Journal of clinical epidemiology[ta]	3,895
#29	Search JAMA[ta]	59,633
#28	Search international journal of epidemiology[ta]	5,682
#27	Search health technology assessment[ta]	484
#26	Search epidemiology[ta]	2,724
#25	Search british medical journal[ta]	107,099
#24	Search annals of internal medicine[ta]	26,014
#23	Search archives of internal medicine[ta]	17,480
#22	Search american journal of preventive medicine[ta]	3,213
#21	Search #15 AND #20	5,576
#20	Search #16 OR #17 OR #18	27,504
#18	Search "Meta-Analysis as Topic"[Mesh]	8,983
#17	Search "Review Literature as Topic"[Mesh]	4,128
#16	Search “systematic review”[tw]	15,722
#15	Search #13 OR #14	432,661
#14	Search #8 OR #10 OR #11 OR #12	355,521
#13	Search #1 OR #2 OR #3 OR #4 OR #5 OR #6	80,379
#12	Search Effect Modifiers (Epidemiology)”[Mesh]	3,276
#11	Search “Research Design/standards”[Mesh]	7,061
#10	Search “Epidemiologic Methods”[Majr]	319,914
#8	Search “Bias (Epidemiology)”[Mesh]	36,384
#5	Search personalized[tw]	3,236
#4	Search tailored[tw]	13,645
#3	Search individualized[tw]	14,183
#2	Search subgroup[tw]	48,785
#1	Search “clinical heterogeneity”[tw]	1,412

Appendix B. Key Informant Questionnaire

Interviewer Name_____

Respondent ID# _____

Publication _____

Phone Number_____

Date and Time_____

Evidence-based Practice Center Systematic Review Protocol
Comparative Effectiveness Methods— Clinical Heterogeneity
Semi-Structured Discussion Guide
0208452.007.009.001

Introduction

Hello (*respondent's name*), this is (*your name*) calling from RTI. I want to thank you for agreeing to talk with me about how researchers have looked for or detected clinical heterogeneity or effect modification during their analyses. To provide you with a little background information about this research study, handling clinical heterogeneity, also referred to as clinical diversity, in systematic evidence reviews (SRs) and comparative effectiveness reviews (CERs) has been an ongoing challenge for Evidence-based Practice Centers (EPCs) and others in developing summary estimates in meta-analyses or narrative assessments. For the purposes of our discussion, we refer to clinical heterogeneity as being a patient characteristic (e.g., age, sex, diagnosis, and disease severity), i.e., factors that cannot be controlled by other means.
The Agency for Healthcare Research and Quality (AHRQ) is interested in learning more about how researchers have considered and dealt with differences in outcomes by patient characteristics via their key questions and in the review process. For this particular task, we are focusing on systematic reviews conducted in the areas of myocardial infarction and osteoarthritis, such as the one that you and your colleagues conducted, to determine how you addressed clinical heterogeneity in the review process.

You were randomly selected to participate in this research study based on your authorship of a systematic review that we accessed through available literature databases entitled, "_____."

As we indicated in our earlier email, our discussion should take approximately 30-45 minutes. Please note that your participation is voluntary and you may refuse to answer any question. Discontinuation will in no way affect any existing relationship that you may have with AHRQ. Because your identity is protected, there is little risk with participation; however, the opportunity for expanding the knowledge base regarding clinical heterogeneity and systematic reviews has a potential benefit.

If you have any questions about this project, you may e-mail or call RTI Project Director, Dr. Suzanne West (swest@rti.org, 919-541-7048) or Dr. Stephanie Chang at AHRQ (_____). If you have any questions about your rights as a study participant you may call RTI's Office of Research Protection toll-free at 1-866-214-2043.

Finally, before we begin, I would like your permission to record our discussion for note-taking purposes. We will destroy the recording when the project is completed. I want to assure you that the information from the discussion that we use to prepare our report to AHRQ will not contain material that can be used to identify you or your institution. Information from this discussion will be aggregated with the responses of the 6-8 other researchers that we interview. [*Await response for a moment, re-ask if needed, and turn on recorder if affirmative.*]

START TIME_____

1. How do you typically approach the development of a study protocol for a systematic review? *[Probe to see if the authors conducted a preliminary search to identify the scope of the literature, talk to experts in the field, etc.]*

   _____
   _____

   1a. Was this the process that you followed for the systematic review titled, *_[insert title of review]*_____?
   Yes…….1 [*Go to 2.*]
   No……..2

   1b. How did the development of this particular study protocol differ from the usual development process that you follow? _____
   _____

We are interested in your opinion on handling clinical heterogeneity, both in developing the key questions for the reviews and analyzing the evidence tables developed from the included publications. The next two questions ask about the formulation of subgroups for a systematic review.

2. What process should be used to identify subgroups for studying clinical conditions in a systematic review? [*Probe for a review of the literature, consultation with experts, etc*]
   _____
   _____
3. At what point in the process should subgroups for study be formulated (e.g., during the protocol development process, or as the information is being extracted from the literature)?
   _____
   _____

This next section discusses how you handled clinical heterogeneity in the review entitled, ___*[insert title of review*]_____ "

4. For this report, we noticed that you did _____. Did you consider any of the following when developing the key questions for this review? *[Read list]*

a. demographic factors	Y	N
b. disease severity (e.g., mild vs. advanced)	Y	N
c. affected joint (*osteoarthritis only*)	Y	N
d. disease recurrence	Y	N
e. other clinical factors [*Specify*: _____]		

[If respondent answered "yes" to any of the above, go to 4a. For all "no" answers, go to 4a2].

4a1. Why did you decide to include ___[*substitute all factors a-e above in which the respondent answered "yes"]*____ in the key questions?

[Probe for using the literature to inform selection, clinical experts, other methods]

	Reason for inclusion
a. demographic factors	_____
b. disease severity (e.g., mild vs. advanced)	_____
c. affected joint (*osteoarthritis only*)	_____
d. disease recurrence	_____
e. other clinical factors	_____

4a2. For those factors above that you chose NOT to include in your key questions, why did you choose to ***exclude*** them?

	Reason for exclusion
a. demographic factors	_____
b. disease severity (e.g., mild vs. advanced)	_____
c. affected joint (*osteoarthritis only*)	_____
d. disease recurrence	_____
e. other clinical factors	_____

5. How did you choose the factors to report on in your analysis?

_____

5a. When, or at what stage did you decide to look for possible clinical heterogeneity due to these factors (e.g., a priori vs a posteriori)?

_____

_____

6. Did you follow any guidance or refer to any manuals during the development of your study protocol?

   Yes…….1 *Specify:* _____ [*Go to 6a*.]
   No……..2 [*Go to 7.]*

   6a. Were you aware of [relevant group – e.g., AHRQ, Centre for Reviews and Dissemination, Cochrane, DERP, IQWiG, NHMRC, NICE, HTA manuals]?

   Yes…….1 [*Go to 6a1*.]
   No……..2 [*Go to 7.]*

   6a1. Did you refer to _____ manual during the development of the study protocol for this review? If so, how was it used?

   _____
   _____

7. Were there any other considerations in your selection of patient or disease factors requiring special consideration? _____
   _____

Those are all the questions that I have for you. Thank you again for your time.

Appendix C. Evidence Tables

Table C1. Guidance documents describing clinical heterogeneity issues

Organization	AHRQ Methods Manual[1]
Year of publication of manual	2009
Does the manual discuss clinical heterogeneity?	Yes (Other terms used: clinical diversity)
How does the manual define clinical heterogeneity?	CH: Variability in study population characteristics, interventions and outcomes. Common examples of factors contributing to CH include: • Age • Sex • Disease severity • Site of lesion • Evolving diagnostic criteria • Change in standard care • Time-dependent care • Difference in baseline risk • Dose-dependent effects
How does the manual define statistical heterogeneity?	SH: Variability in the observed treatment effects being evaluated in different trials.
Does the manual discuss the relationship between clinical and statistical heterogeneity?	Diversity in clinical characteristics will cause SH if the true treatment effect varies depending on those characteristics.
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	No recommendation
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion
What is the recommendation for addressing clinical heterogeneity in a systematic review?	• Determine factors expected to account for clinical heterogeneity a priori based on good knowledge of the clinical and biological background of the topic from previous reviews or experts • Frame key questions considering factors contributing to CH, including subgroup analyses as necessary • When conducting meta-analyses in the presence of CH, use meta-regression with sensitivity analyses. • Define the threshold for acceptable differences in clinical characteristics which could be combined in a meta-analysis based on the scope of the research question. • Provide rationale for deciding whether to combine studies when CH present

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

What is the recommendation for addressing statistical heterogeneity in a systematic review?	Explore statistical heterogeneity (e.g., meta-regression, control rate meta-regression, subgroup analysis). With substantial unexplained heterogeneity, one cannot (or at least should not) determine a precise estimate of treatment effect, but one may still be confident in the direction of effect.
Other comments	
Organization	Centre for Reviews and Dissemination's (CRD) Systematic Reviews: Guidance for undertaking reviews in health care[2]
Year of publication of manual	2009
Does the manual discuss clinical heterogeneity?	Yes (Other terms used: clinical diversity)
How does the manual define clinical heterogeneity?	CH: differences in participants, interventions or outcome measures
How does the manual define statistical heterogeneity?	SH: differences in the effect estimates; variation other than that which arises by chance; reflects methodological or clinical differences between studies
Does the manual discuss the relationship between clinical and statistical heterogeneity?	If clinical diversity influences the estimated intervention effect, there will be some statistical heterogeneity between studies.
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	• Visually examine forest plots • Chi-squared tests (Q-statistic) • I^2 test (percentage of variability in the effect estimate attributable to heterogeneity rather than chance)
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	• Any specified restrictions should be clinically justifiable and relevant • The included population should be relevant to the population to which the review findings will be applied
What is the recommendation for addressing clinical heterogeneity in a systematic review?	• Sources of CH can be explored using subgroup analyses; ideally, subgroups should be planned at the protocol stage. • Where subgroup analysis is dependent upon what data are available, and an adaptive process is needed, this should be stated clearly in the protocol.
What is the recommendation for addressing statistical heterogeneity in a systematic review?	If intervention effects vary with different populations or intervention characteristics consider subgroup analyses or metaregression.
Other comments	CH information available only in section on "Systematic Reviews of Clinical Tests"; not addressed specifically in other sections.
Organization	Cochrane Handbook for Systematic Reviews of Interventions[3] From: http://www.cochrane-handbook.org/ (Part II: Sections 9.5 & 9.6)
Year of publication of manual	2008
Does the manual discuss clinical heterogeneity?	Yes (Other terms used: clinical diversity)
How does the manual define clinical heterogeneity?	CH: Variability in the participants, interventions and outcomes studied
How does the manual define statistical heterogeneity?	SH: Variability in the observed intervention effects being evaluated in the different studies beyond what one would expect from chance alone (random error); a consequence of clinical or methodological diversity (variability in study design or risk of bias), or both, among the studies. Statistical heterogeneity manifests itself in observed intervention effects being more different from each other than one would expect from chance alone (random error).

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

Does the manual discuss the relationship between clinical and statistical heterogeneity?	Clinical variation will lead to heterogeneity if the intervention effect is affected by the factors that vary across studies; most obviously, specific interventions or patient characteristics. In other words, the true intervention effect will be different in different studies.
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	• Consider the overlap of confidence intervals in forest plots • Consider formal statistical testing: chi-squared; because power is low, a P of 0.10 is recommended. P-value can provide the strength of evidence for I-squared • Consider meta-regression • Ideally, pre-specify investigations of characteristics of studies that may be associated with heterogeneity in the protocol. • When writing the analysis section of the protocol, consider how clinical diversity will be assessed and whether (and how) it will be incorporated into the analysis strategy. • Consider quantifying inconsistency across studies, moving focus away from testing for heterogeneity to assessing its impact on the meta-analysis (e.g., using I^2 statistic). The importance of the observed value depends on the magnitude and direction of effect and the strength of the evidence.
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	• Any restrictions with respect to specific population characteristics or settings should be based on a sound rationale. • It is important that reviews are globally relevant, so justification for the exclusion of studies based on population characteristics should be explained in the review. • When it is uncertain whether there are important differences in effects among various subgroups of people, it may be best to include all of the relevant subgroups and then test for important and plausible differences in effect in the analysis
What is the recommendation for addressing clinical heterogeneity in a systematic review?	• If intervention effects vary with different populations or intervention characteristics consider subgroup analyses or metaregression. • Only consider meta-analysis when studies are sufficiently homogeneous in terms of participants, interventions and outcomes to provide a meaningful summary.
What is the recommendation for addressing statistical heterogeneity in a systematic review?	• Check again that data is correct • Don't do meta-analysis • Explore heterogeneity (subgroup analyses or meta-regression) • Change the effect measure • Perform random effects meta-analysis Since clinical and methodological diversity always occur in a meta-analysis, heterogeneity always exists whether or not it can be detected using a statistical test.
Other comments	Cochrane distinguishes between qualitative (reversed direction) and quantitative (same direction, difference in magnitude of effects) of effect modification. From http://www.cochrane-net.org/openlearning/HTML/mod13-2.htm • The term 'clinical diversity' (sometimes called 'clinical heterogeneity') describes clinical differences in the studies to do with the participants, interventions and outcomes. • study location and setting • age, sex, diagnosis and disease severity of participants • treatments people may be receiving at the start of a study • dose or intensity of the intervention • definitions of outcomes.
Organization	Drug Effectiveness Review Project (DERP II)[4] From: Review Methods and Report Production Procedures

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

Year of publication of manual	Rev. 1, April 2008
Does the manual discuss clinical heterogeneity?	No
How does the manual define clinical heterogeneity?	Heterogeneity: The variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies. (also called quantitative heterogeneity; the term qualitative heterogeneity is used, but never defined)
How does the manual define statistical heterogeneity?	CH and SH are not distinguished explicitly.
Does the manual discuss the relationship between clinical and statistical heterogeneity?	CH and SH are not distinguished explicitly.
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	• In formulating key questions, consider if comparative effectiveness, or tolerability and safety vary in patient subgroups by demographics (age, racial groups, gender, etc), use of other medications, or presence of co-morbidities
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	• No discussion
What is the recommendation for addressing clinical heterogeneity in a systematic review?	(To deal with general heterogeneity) • Study quality and qualitative heterogeneity across studies in study design, patient population, interventions, and outcomes, are considered in order to determine whether meta-analysis should be performed. • If meta-analysis cannot be performed, summarize the data qualitatively.
What is the recommendation for addressing statistical heterogeneity in a systematic review?	No recommendation
Other comments	None
Organization	HuGE Net[5] From: The HuGE Net HuGE Review Handbook, Version 1.0
Year of publication of manual	
Does the manual discuss clinical heterogeneity?	No; discusses general "heterogeneity" and defines it as *variation in associations across studies*
How does the manual define clinical heterogeneity?	Not defined
How does the manual define statistical heterogeneity?	Not defined
Does the manual discuss the relationship between clinical and statistical heterogeneity?	No
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	N/A; To assess general heterogeneity: • estimate of among-study variance/I^2 statistic • sensitivity analyses • meta-regression • cumulative meta-analysis • recursive cumulative meta-analysis
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

What is the recommendation for addressing clinical heterogeneity in a systematic review?	For individual studies where subgroup analyses have been reported by disease or socio-demographic characteristics: • indicate what subgroups have been analyzed (e.g., subsite of tumor, tumor histology, age, ethnic group), • discuss results in the text rather than trying to summarize all subgroup analyses in a table, unless subgroups form one of the main pre-specified analyses of the systematic review.
What is the recommendation for addressing statistical heterogeneity in a systematic review?	No recommendation
Other comments	None
Organization	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)[6] (English: Institute for Quality and Efficiency in Health Care)
Year of publication of manual	
Does the manual discuss clinical heterogeneity?	No; discusses general "heterogeneity" and defines it as large differences shown between the results of individual studies.
How does the manual define clinical heterogeneity?	Not defined
How does the manual define statistical heterogeneity?	Not defined
Does the manual discuss the relationship between clinical and statistical heterogeneity?	No
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	N/A; To assess general heterogeneity: • If a priori information is available on a possible effect modifier (e.g. age, pathology), investigate possible heterogeneity in advance with regard to the effect in the various patient groups. • Consider information on a possible heterogeneity of patient groups. appropriately in the study design • If necessary, conduct several studies. • Only perform a meta-analytical summary of strongly heterogeneous study results if the reasons for this heterogeneity are plausible and still justify such a summary • quantify impact of heterogeneity (I^2) • Use meta-regression
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion
What is the recommendation for addressing clinical heterogeneity in a systematic review?	As a minimal prerequisite for the Institute to use a SR on the effects of treatments, consider whether the results have been consistent among different populations and subgroups (e.g., gender and baseline disease risk).
What is the recommendation for addressing statistical heterogeneity in a systematic review?	No recommendation
Other comments	None
Organization	National Health and Medical Research Council (NHMRC)[7] Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

Year of publication of manual	
Does the manual discuss clinical heterogeneity?	No; discusses "heterogeneity" and defines it as: 1) when results vary among the studies more than can be attributed to chance; 2) the differences in treatment effect between studies contributing to a meta-analysis.
How does the manual define clinical heterogeneity?	Not defined
How does the manual define statistical heterogeneity?	The term "statistical heterogeneity" is used but not defined.
Does the manual discuss the relationship between clinical and statistical heterogeneity?	No; Guide does state: Difference in the effects seen may be caused by several factors, including disease features such as stage or severity.
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	N/A; To assess general heterogeneity: • Even if test for heterogeneity is nonsignificant, exploring for causes of variation is reasonable and useful (Nothing more specific regarding handling of potential heterogeneity is discussed in detail.)
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion
What is the recommendation for addressing clinical heterogeneity in a systematic review?	No recommendation
What is the recommendation for addressing statistical heterogeneity in a systematic review?	No recommendation
Other comments	None
Organization	National Institute for Health and Clinical Excellence (NICE)[8] From: The Guidelines Manual & Appendices
Year of publication of manual	JAN 2009
Does the manual discuss clinical heterogeneity?	No; discusses "heterogeneity" and defines it as: when results or effect estimates of treatment from separate studies seem to be very different; may occur because of differences between studies in patient populations, outcome measures or definition of variables.
How does the manual define clinical heterogeneity?	Not defined
How does the manual define statistical heterogeneity?	Not defined
Does the manual discuss the relationship between clinical and statistical heterogeneity?	No
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	N/A; To assess general heterogeneity: Describe and justify meta-analytical techniques and approaches to dealing with heterogeneity, including specification of any subgroup analyses and sensitivity analyses.
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion

Table C1. Guidance documents describing clinical heterogeneity issues (continued)

What is the recommendation for addressing clinical heterogeneity in a systematic review?	No recommendation
What is the recommendation for addressing statistical heterogeneity in a systematic review?	No recommendation
Other comments	None
Organization	European Network for Health Technology Assessment (EUNetHTA)[9]
Year of publication of manual	2008
Does the manual discuss clinical heterogeneity?	Yes
How does the manual define clinical heterogeneity?	**Clinical heterogeneity**: differences in participant characteristics, interventions, and outcome measures
How does the manual define statistical heterogeneity?	Not defined
Does the manual discuss the relationship between clinical and statistical heterogeneity?	No
What is the recommendation for assessing whether clinical heterogeneity affects the intervention effect?	No recommendation
Does the manual discuss restriction as a way of addressing clinical heterogeneity?	No discussion
What is the recommendation for addressing clinical heterogeneity in a systematic review?	• Tables are useful to describe populations, interventions, settings, outcome measures • Provide statements about the presence of clinical heterogeneity
What is the recommendation for addressing statistical heterogeneity in a systematic review?	• Explore heterogeneity and consider in discussion • Perform sensitivity analyses based on clinical issues • Perform random effects meta-analysis
Other comments	None

References

1. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Available at http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Accessed 2007.
2. Centre for Reviews and Dissemination. Centre for Reviews and Disssemination's (CRD) Systematic Reviews: Guidance for undertaking reviews in health care. Available at http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf. Accessed March 9, 2009.
3. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1. Available at http:www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf. Accessed 2008.

4. Project DER. Review methods and report production procedures. Available at http://www.ohsu.edu/ohsuedu/researchpolicycenter/DERP/. Accessed May 28, 2009.
5. The HuGENet Writing Group. The HuGENET™ HuGE Review Handbook, Version 1.0. Available at http://www.hugenet.ca. Accessed February 28, 2009.
6. Gesundheitswesen IfQuWi. General methods. Available at http://www.iqwig.de/downloadIQWiG_General_methods_V-3-0.pdf. Accessed October 19, 2009.
7. National Health and Medical Research Council. How to review the evidence: systematic identification and review of the scientific literature. Available at http://www.nhmrc.gov.au/_files_nhmrc/file_publications/synopses/cp65.pdf. Accessed October 9, 2009.
8. National Institute for Health and Clinical Excellence. The guidelines manual. Available at http://www.nice.org.uk/media/5F2/44_The_guidelines_manual_2009_-_All_chapters.pdf. Accessed October 19, 2009.
9. European Network for Health Technology Assessment. HTA core model for medical and surgical interventions 1.0R. Available at http://www.eunethta.net/upload/WP4/Final%20Deliverables/HTA%20Core%20Model%20for%20Medical%20and%20Surgical%20Interventions%201%200r.pdf. Accessed March 9, 2009.

Abbreviations

3D-CRT	Three-dimensional ***conformal radiotherapy***
ACC	American College of Cardiology
ACE-1	angiotensin-converting enzyme inhibitor
ACR	American College of Radiology or American College or Rheumatology
ACS	***American Cancer Society***
AD	antidepresant
AF	Atrial Fibrillation
AHA	American Heart Association
AIDS	***Acquired immune deficiency syndrome***
AiiRA	angiotensin II receptor antagonist
AL-TENS	acupuncture-like transcutaneous nerve stimulation
AMED	Allied and Complimentary Medicine Database
AMI	acute myocardial infarction
ARBs	angiotensin receptor blockers
B	beta
BCS	British Cardiac Society
BIRADS	Breast Imaging Reporting and Data System
BMI	body mass index
BP	blood pressure
CAD	coronary artery disease
CAE-1	coronary artery ectasia
CBT	Cognitive Behavioral Therapy
CCBs	calcium channel blockers
CDMR	caesarean delivery on maternal request
CH	clinical heterogeneity
CHF	congestive heart failure
CKD	chronic kidney disease
cm	centimeter
CMF	cyclophosphamide, methotrexate, and fluorouracil
CMH	Chinese medicinal herbs
COPD	chronic obstructive pulmonary disease
COX-2	Cycloosygenase-2 (inhibitors)
CR	complete remission
CT	Computed Tomography
CV	cardiovascular
DCIS	ductal carcinoma in situ
DERP	Drug Effective Review Project
DM	diabetes mellitus
DRI	Diabetes Research Institute
DSM-IV	Diagnostic Statistical Manual of Mental Disorders, IV edition
E.S.	executive summary
EBB	endobronchial biopsies
EBRT	external brain radiation therapy
ECG	electrocardiogram

ECT	electro convulsive therapy
ECTCG	Early Breast Cancer Trialists' Collaborative Group
ED-SCLC	extensive disease - small cell lung cancer
EFW	estimated fetal weight
ER	emergency room
ES	electrical stimulation
EUS-FNA	endoscopic ultrasound or ultrasonograph – fine needle aspiration
EWS	early warning symptoms and signs
FEV_1	forced expiratory volume
FL	frontal lobe; fetal length
FM	fetal movement
FNA	fine needle aspiration
GDG	Guideline Development Group
GERD	gastroesophageal reflux disorder
GI	gastrointestinal
GNRH	gonadotrophin-releasing hormone
Gp	glycoprotein
GPs	general practitioners
Gv	gastric volume
HA	headache; hepatitis A; hospital admission
HDL-c	high density lipoprotein-cholesterol
HER2	human epidermal growth factor receptor 2
HF	heart failure
HIV	human immunnodefiency virus
HLD	herniated lumbar disk
HTN	hypertension
IAS	infant apnea syndrome; interarterial shunt
IBS	irritable bowel syndrome
ICD	International Classification of Diseases
ICD-10	International Classification of Diseases - 10
ICS	inhaled corticosteroid
IMRT	intensity-modulated radiation therapy
IUS	intraopreative ultronosonography
KCQs	key clinical questions
KL	Kellgren-Lawrence (radiographic)
KQ	key question
LDCT	low-dose computed tomography
LDL-c	low density lipoprotein-cholesterol
LD-SCLC	limited state-small cell lung cancer
LES	lower esophageal sphincter (or stricture)
LHRH	luteinizing hormone releasing hormone
LMWH	low molecular weight heparin
LV	left ventricle
LVEF	left ventricle ejection fraction
LVSD	left ventricle systolic dysfunction (or diameter)
MBL	menstrual blood loss

MDD	major depressive disorder
mg/dL	milligrams per deciliter
MI	myocardial infarctin
ml	milileter
mmHg	millimeter of mercury
mmol	millimoles
mmol/1	millimoles per liter
MRI	magnetic resonance imaging
N2/N3	ipsilateral or subcarinal nodes/contralateral mediastinal nodes
NA	not applicable
NPPV	noninvasive positive pressure ventilator
NSAIDs	non-steroidal nasal inflammatory drug
NSCLC	non-small cell lung cancer
NYHA	New York Hospital Association
OA	osteoarthritis
OGTT	Oral Glucose Tolerance Test
P	probability
PAH	pulmonary arterial hypertension
PCI	prophylactic cranial irradiation
PET	positron emission tomography
PICO	population intervention comparator outcomes
PPCI	pediatric pain coping inventory
PPD	postpartum depression
PPIs	proton pump inhibitor
PSA	prostate specific antigen
pt	patient
RAS	rennin angiotensin system
RCTs	randomized controlled trials
RDC	Research Diagnostic Criteria
RFA	radiofrequency ablation
RP	radial pulse; radical prostectomy; rising pressure
RRT	radical radiation therapy; registered recreation therapist; renal replacement therapies
RT	radiation therapy; response time
rTMS	repetitive transcranial magnetic therapy
SAD	seasonal affective disorder; schizo affective disorder
SADs	schedule for affective disorders and schizophrenia; social anxiety and distress scale
SC	scintimammography; scan; sciatia; self-care; spinal cord- sigmoid colon; subcortical; subcutaneous
SCLC	small cell lunc ancer
SEGT	subependymal giant cell tumor
SLNB	sentinel lymph node biopsy
SNRI	serotonin noradrenaline reuptake inhibitor
SSNS	sick sinus node syndrome
SSRI(s)	selective serotonin reuptake inhibitor(s)

STEMI	ST-elevation myocardial infarction
TAU	treatment as usual
TCM	Traditional Chinese Medicine
TENS	transcutaneous electrical nerve stimulation
TNS	transcutaneous nerve stimulation
TRTx	tratment
TZDs	thiazolidinedione derivatives
U.S.	United States
UFH	uterine funal height; unfractionated heparin
UK	United Kingdom
US	ultrasonography
vs.	versus
WLE	wide local excision

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, BREAST				
Bruening et al., 2006[1]	KQ1: For the following diagnostic tests as applied to the breast (PET scanning, SC, MR, and US) what are the sensitivity and specificity of the tests for diagnosis of breast cancer in women presenting with: a. An abnormal mammogram, overall and by BIRADS classification or other relevant clinical classification (e.g., presence or absence of calcification, well circumscribed lesions, etc.) b. A palpable breast abnormality c. What percentage of women in the studies in this question were age 65 or older, and do sensitivity and specificity vary by older vs. younger than age 65?	• Age	Not stated explicitly	Note: Authors report that evidence bases were unacceptably weak for evaluation of test performance for many subgroup comparisons. • Age (older vs. younger than age 65) • Pre- vs. post-menopausal • Suspicious vs. palpable lesion • Breast tissue (dense vs. fatty) • BIRADS status • Morphological characteristics of lesion (e.g., size, presence or absence of calcifications, etc.)

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: For women with relevant demographic risk factors (e.g., age, family history) and clinical risk factors (e.g., BIRADS status or morphologic characteristics of the lesion), what are the positive and negative predictive values of the above diagnostic tests?	• Risk factors • Age	Not stated explicitly	Note: Authors report that evidence bases were unacceptably weak for evaluation of test performance for many subgroup comparisons. • Women with relevant demographic and clinical risk factors • Age (older vs. younger than age 65) • Pre- vs. post-menopausal • Suspicious vs. palpable lesion • Breast tissue (dense vs. fatty) • BIRADS status • Morphological characteristics of lesion (e.g., size, presence or absence of calcifications, etc.)
	KQ3: Are there other factors that affect the accuracy or acceptability of the tests considered in Question 1 and 2?	No	NA	NA

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, LUNG				
Seidenfeld et al., 2006[2]	KQ1: For limited-stage SCLC, what are the relative benefits and harms (survival, toxicity, and quality of life) of TRTx combined with chemotherapy, either in alternating fashion, concurrently or sequentially?	• Severity, stage, or site	Not stated explicitly	• Patients with a histopathologically confirmed diagnosis of SCLC staged as limited disease
	KQ2: For limited-stage SCLC, do outcomes (survival, toxicity, or quality of life) differ if concurrent TRTx is given in early vs. late chemotherapy cycles?	• Severity, stage, or site	Not stated explicitly	• Patients with a histopathologically confirmed diagnosis of SCLC staged as limited disease
	KQ3: For limited-stage SCLC, do outcomes (survival, toxicity, quality of life) of primary therapy differ if one varies dose rate, treatment interval, or fractionation scheme for delivering TRTx? Comparisons of interest include: • accelerated regimens (>10 Gv per week completed over a short interval) vs. standard duration regimens (≥ Gv per week) vs. split courses delivered over the standard interval; and • single daily fractions vs. hyperfractionated (two or more daily fractions or concomitant boost).	• Severity, stage, or site	Not stated explicitly	• Patients with a histopathologically confirmed diagnosis of SCLC staged as limited disease
	KQ4: What are the relative benefits and harms (survival, toxicity, and quality of life) of adding TRTx to chemotherapy for primary treatment of extensive-stage SCLC?	• Severity, stage, or site	Not stated explicitly	• Patients with histopathologically confirmed diagnosis of SCLC staged as extensive disease undergoing first-line therapy

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ5: What are the benefits and harms (survival, toxicity and quality of life) of PCI for patients with SCLC in CR after primary therapy?	• Severity, stage, or site	Not stated explicitly	• Patients with histopathologically confirmed diagnosis of SCLC that has completely responded to primary therapy (regardless of stage)
	KQ6: Does the addition of PET scanning improve the accuracy of staging for patients diagnosed with SCLC, over the use of other techniques, including CT and MRI, without PET?	• Co-interventions	Not stated explicitly	• Patients with histopathologically confirmed diagnosis of SCLC
	KQ7: What are the outcomes (survival, toxicity and quality of life) of treatments used to manage patients with mixed small cell/non-small cell lung cancers?	• Comorbidities	Not stated explicitly	• Patients with a histopathologically confirmed diagnosis of mixed small cell/non-small cell lung cancer
	KQ8: What is the role of surgery and what is its impact on survival in patients with very early stage SCLC? How do available studies define very early stage SCLC?	• Severity, stage, or site	Not stated explicitly	• Patients with histopathologically confirmed diagnosis of SCLC staged as limited disease with small tumors and no nodal involvement

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ9: What are the outcomes of second- or subsequent-line therapy in patients with relapsed or progressive SCLC? Where available data permit, patients with limited- and extensive-stage disease will be addressed separately, as will those with refractory disease (relapse or progression within 3 months of primary treatment).	• Severity, stage, or site	Not stated explicitly	• Patients with histopathologically confirmed diagnosis of SCLC that either relapsed or progressed after a response that lasted at least 3 months following primary therapy for: (a) limited-stage or (b) extensive-stage disease; or (c) patients with refractory disease (defined as no response or progression within 3 months of primary therapy)
CANCER, PROSTATE				
Wilt et al., 2008[3]	KQ1: What are the comparative risks, benefits, and outcomes of therapies for clinically localized prostate cancer?	No	NA	NA
	KQ2: How do patient characteristics, e.g., age, race/ethnicity, presence or absence of comorbid illness, preferences (e.g., tradeoff of treatment-related adverse effects vs. potential for disease progression), affect the outcomes of these therapies, overall and differentially?	• Age, race/ethnicity • Comorbidities	Not stated explicitly	• Race/ethnicity • Age • Comorbid illness
	KQ3: How do provider/hospital characteristics affect outcomes overall and differentially (e.g., geographic region and volume)?	No		

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ4: How do tumor characteristics, e.g., Gleason score, tumor volume, screen vs. clinically detected tumors, affect the outcomes of these therapies, overall and differentially?	• Severity, stage, or site	Not stated explicitly	• Authors noted that not enough data existed based on PSA levels, histologic score, and tumor volume to identify low-, intermediate-, and high-risk tumors so authors focused on PSA levels and Gleason histological scores • Age was looked at in the analysis
HEART FAILURE				
Shekelle et al., 2003[4]	KQ1: What evidence exists on the effectiveness of nurse management programs? Health food supplements?	No	NA	NA
	KQ2: What evidence exists on the treatment of sleep apnea in patients with HF?	No	NA	NA
	KQ3: What is the evidence on the treatment of specific myocardial disorders; e.g., myocarditis, sarcoidosis, and amyloidosis, in patients with HF?	No	NA	NA
	KQ4: What interventions are effective for patients with diastolic dysfunction?	No	NA	NA
	KQ5: Which patients benefit from which beta-blockers?	No	NA	NA
	KQ6: What are the effects of potassium levels on HF outcomes?	No	NA	NA
	KQ7: Do angiotensin blockers improve outcomes?	No	NA	NA
	KQ8: What, if any, are the differences in treatment effectiveness associated with patient gender, race, age, and income level?	• Age, race, sex, income level	Not stated explicitly	*This question was not used in final analysis. See revised KQs below.*

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	The AHA and the ACC released practice guidelines on the management of HF, so KQs for this report were revised in order to compliment, rather than duplicate, the AHA/ACC report. After consulting a technical expert panel, the following were considered areas in which significant contribution could still be made.			
	• Assessment of the effects of age over 70, gender, race, and assisted living on treatment outcomes • Cost-effectiveness of medication combinations • Assessment of outcomes in patients with various comorbidities, particularly diabetes mellitus, renal dysfunction, and cognitive dysfunction	• Age, race, sex, assisted living • Comorbidities	Not stated explicitly	• Age (>70 years) • Gender • Race (black vs. white) • Comorbidities (diabetes mellitus, renal dysfunction, cognitive dysfunction)
CESAREAN SECTION				
Viswanathan et al., 2006[5]	KQ1: First, what is the trend and incidence of cesarean delivery over time in the United States and in other developed countries? Secondarily: • What is the contribution of primary prelabor cesarean deliveries? • Of the primary prelabor cesarean deliveries, what is the contribution of CDMR, for medical indications, and for malpresentation?	No	NA	NA
	KQ2:Outcomes of Cesarean Delivery on Maternal Request	No	NA	NA
	KQ3: What factors affect the magnitude of the benefits and harms identified in KQ2?	• Race/ethnicity, sex (fetal), socioeconomics • Comorbidities • Risk Factors	No	• Maternal age • Gestational age • Pre-pregnancy BMI • Race • Physician experience • Incision type • Prophylactic antibiotics

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CHRONIC OBSTRUCTIVE PULMONARY DISEASE				
Wilt et al., 2005[6]	KQ1: What is the prevalence of COPD and airflow obstructions in various adult populations as defined by: (1) spirometry and (2) clinical examination?	No	NA	
	KQ2: Can use of spirometry lead to increased smoking cessation rates?	No	NA	Looked at only in analysis: • Gender • Race • Smoking intensity (reported in pack-years) • Number of previous attempts to quit, if any
	KQ3: Does the effectiveness of specific therapies to improve clinically relevant outcomes in COPD vary based on baseline or follow-up spirometry, short-term spirometric response due to initial therapy, or spirometric progression over time?	No	NA	NA
	KQ4: Is prediction of prognosis based on spirometry, with or without clinical indicators, more accurate than prognosis based on clinical indicators alone?	No	NA	NA
DEPRESSION				
Gartlehner et al., 2007[7]	KQ1a: For adults with MDD, dysthymia, or subsyndromal depressive disorders, do commonly used medications for depression differ in efficacy or effectiveness in treating depressive symptoms?	• Age	Not stated explicitly	• Adult inpatients and outpatients with MDD, dysthymia, of subsyndromal depression
	KQ1b: If a patient has responded to one agent in the past, is that agent better than current alternatives at treating depressive symptoms?	• Severity, stage, or site		• Adult inpatients and outpatients with MDD, dysthymia, of subsyndromal depression

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2a: For adults with a depressive syndrome, do antidepressants differ in their efficacy or effectiveness for maintaining response or remission (i.e., preventing relapse or recurrence)?	No		• Adult inpatients and outpatients with a history of depressive illnesses currently in remission
	KQ2b: For adults receiving antidepressant treatment for a depressive syndrome that either has not responded (acute phase) or has relapsed (continuation phase) or recurred (maintenance phase), do alternative antidepressants differ in their efficacy or effectiveness?	• Age • Severity, stage, or site		• Adult inpatients and outpatients with recurrent depression • Subgroup analyses: • Response to treatment (no response, relapse or recurrence)
	KQ3: Do medications or combinations of medications (including tricyclics in combination) used to treat depression differ in their efficacy or effectiveness for treating accompanying symptoms, such as anxiety, insomnia, and neurovegetative symptoms? 3a: Do medications differ in their efficacy and effectiveness in treating the depressive episode? 3b: Do medications differ in their efficacy and effectiveness in treating accompanying symptoms?	No	NA	• Adult inpatients and outpatients with MDD, dysthymia, of subsyndromal depression
	KQ4: For adults with a depressive syndrome, do commonly used antidepressants differ in safety, adverse events, or adherence? Adverse effects of interest include but are not limited to nausea, diarrhea, headache, tremor, daytime sedation, decreased libido, failure to achieve orgasm, nervousness, insomnia, and more sever events including suicide.	No	NA	• Adult inpatients and outpatients with MDD, dysthymia, or subsyndromal depression

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ5: How do the efficacy, effectiveness, or harms of treatment with antidepressants for a depressive syndrome differ for the following subpopulations: • Elderly or very elderly patients; • Other demographic groups (defined by age, ethnic or racial groups, and sex); • Patients with medical comorbidities (e.g., ischemic heart disease, cancer) • Patients with psychiatric and behavioral comorbidities (e.g., substance abuse disorders; and • Patients taking other medications.	• Age, race/ethnicity, sex • Co-interventions • Comorbidities	Not stated explicitly	• Age (in general, and specifically elderly or very elderly) • Race or ethnicity • Sex • Comorbidities, medical (e.g., ischemic heart disease, cancer) or psych/behavioral (e.g., SADs) • Concurrent medications Defined in analysis as: • Adult inpatients and outpatients with MDD, dysthymia, of subsyndromal depression
DYSPEPSIA	AHRQ has not produced any report specifically for dyspepsia. We used a report GERD instead.			
Ip et al., 2005[8]	KQ1A: What is the evidence of the comparative effectiveness of medical, surgical, and endoscopic, treatments in improving objective and subjective outcomes in patients with chronic GERD?	No	NA	NA
	KQ1B: In patients with Barrett's esophagus, what is the result of medical vs. surgical management in terms of the incidence of adenocarcinoma of the esophagus?	• Risk factor	NA	NA

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2A: What are the characteristics of patients who have undergone these therapies, including the nature of previous medical therapy, severity of symptoms, age, sex, weight, other demographic and medical factors or by specific patient subgroups, and provider characteristics for procedures including provider volume and setting (e.g., academic vs. community)?	No	NA	NA
	KQ2B: Is there evidence that effectiveness of medication, surgical and endoscopic treatments vary for specific patient subgroups?	• Unspecified subgroups	Not stated explicitly	• Age • Sex • BMI • Psychological profile • Baseline symptoms • Preoperative response to acid-suppression therapy • Esophagitis (any severity) • Esophagitis (grade 3 or 4) • Severity of acid reflux • LES competence • LES pressure • Esophageal motility • Hiatal hernia
	KQ3: What are the short- and long-term adverse effects associated with specific medical, surgical, and endoscopic therapies for GERD? Does the incidence of adverse effects vary with duration of follow-up, specific surgical intervention, or patient characteristics?	• Demographics (unspecified) • Co-interventions	Not stated explicitly	• Age • Severity of GERD • Presence of *H. pylori* • Baseline symptoms

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HYPERTENSION				
Matchar et al, 2007[9]	KQ1: For adult patients with essential hypertension, how do ACEIs and ARBs differ in blood pressure control, cardiovascular risk reduction, cardiovascular events, quality of life, and other outcomes?	No	NA	• Age (adults 18+ years) • Diagnosis of essential hypertension (as defined by study authors)
	KQ2: For adult patients with essential hypertension, how do ACEIs and ARBs differ in safety, adverse events, tolerability, persistence, and adherence?	No	NA	• Age (adults 18+ years) • Diagnosis of essential hypertension (as defined by study authors)
	KQ3: Are there subgroups of patients based on demographic characteristics (age, racial and ethnic groups, sex), use of other medications concurrently, or comorbidities for which ACEIs or ARBs are more effective, associated with fewer adverse events, or better tolerated?	• Age, race/ethnicity, sex • Co-interventions (medications) • Comorbidities	Not stated explicitly	• Demographics (age, racial and ethnic groups, sex) • Concurrent medication use • Comorbidities
LABOR INDUCTION				
Caughey et al., 2009[10]	KQ1: What evidence describes the maternal risks of elective induction vs. expectant management?	No	NA	NA
	KQ2: What evidence describes the fetal/neonatal risks of elective induction vs. expectant management?	No	NA	NA
	KQ3: What is the evidence that certain physical conditions/patient characteristics (e.g., parity, cervical dilatation, previous pregnancy outcome) are predictive of a successful induction of labor?	• Risk factors	Not stated explicitly	• Parity • Cervical Status • Maternal age • Maternal BMI • Gestational age • Amniotic fluid index
	KQ4: Definition of Successful Labor Induction	No	NA	NA

Table C2. Agency for Healthcare Research and Quality (AHRQ) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
OSTEOARTHRITIS				
Chou et al., 2006[11]	KQ1a: What are the comparative benefits and harms of treating osteoarthritis with oral medications or supplements?	No	NA	NA
	KQ1b: How do these benefits and harms change with dosage and duration of treatment, and what is the evidence that alternative dosage strategies, such as intermittent dosing and drug holidays, affect the benefits and harms of oral medication use?	No	NA	NA
	KQ2: Do the comparative benefits and harms of oral treatments for osteoarthritis vary for certain demographic and clinical subgroups? Demographic subgroups (age, sex, and race) Coexisting diseases (hypertension, edema, ischemic heart disease, heart failure; peptic ulcer disease; history of previous bleeding due to NSAIDs) Concomitant medication use includes anticoagulants	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, sex, and race) • Co-existing diseases (hypertension, edema, ischemic heart disease, heart failure, peptic ulcer disease; history of previous bleeding due to NSAIDs) • Concomitant anticoagulant or aspirin use
	KQ3: What are the comparative effects of co-prescribing of H2-antagonists, misoprostol, or PPIs on the gastrointestinal harms associated with NSAID use?	No	NA	NA
	KQ4: What are the comparative benefits and harms of treating osteoarthritis with oral medications as compared with topical preparations?	No	NA	NA

Evidence Table C3. Cochrane final reports

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, BREAST				
Zhang et al., 2007[12]	Objective: To assess the effectiveness and safety of CMH in alleviating chemotherapy-induced short term side effects in breast cancer patients.	No	NA	Planned to carry out the following subgroup analyses: • Age (perimenopausal, postmenopausal) • Type of primary treatment • Early compared to advanced breast cancer • Participants receiving chemotherapy drugs, different durations of treatment and different dosages • Duration of follow up: on the basis of data • Unable to perform subgroup analyses owing to the small number of trials. Inclusion Criteria: • Female breast cancer patients receiving chemotherapy either as adjuvant treatment for early or advanced breast cancer or as palliative treatment for metastatic breast cancer and experiencing chemotherapy-induced side effects.

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Early Breast Cancer Trialists' Collaborative Group (EBCTCG)[13]	Objective: Assess the 10-year and 15-year effects of various systemic adjuvant therapies on breast cancer recurrence and survival by comparing: • single-agent chemotherapy vs. no adjuvant chemotherapy • polychemotherapy vs. no adjuvant chemotherapy • anthracycline-based polychemotherapy vs. standard polychemotherapy with CMF • longer vs. shorter polychemotherapy • tamoxifen vs. no adjuvant tamoxifen • longer vs. shorter tamoxifen; or • ovarian ablation or suppression (in women of age <50 years) vs. no adjuvant ovarian treatment	No	NA	• Age • Type of polychemotherapy regimen • Presence or absence of tamoxifen in both treatment groups • ER status and tamoxifen use • Nodal status • Period of follow-up

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, LUNG				
Amarasena et al., 2008[14]	Objectives: To determine the effectiveness of platinum chemotherapy regimens compared with non-platinum chemotherapy regimens in the treatment of SCLC with respect to survival, tumour response, toxicity and quality of life. We will undertake the following comparison: • platinum agents vs. other chemotherapeutic agents [P vs. A]; • platinum agents combined with other chemotherapy agents vs. the same chemotherapy regimen without the platinum agents [(P+A) vs. A]; • platinum agents combined with other chemotherapy agents vs. any other chemotherapy regimens without platinum agents [(P+A) vs. B]. (Where P = platinum chemotherapy agents, A = non-platinum chemotherapy regimens and B = non-platinum chemotherapy regimens [different from A]).	No	NA	• Disease stages (Undifferentiated, LD-SCLC, or ED-SCLC) Selection Criteria: • Patients with pathologically confirmed (cytological or histological) SCLC
Cadona Zorilla et al., 2008[15]	Objective: To assess the effectiveness of palliative EBB in increasing survival and to control thoracic symptoms in patients with advanced NSCLC compared with EBRT or other alternative endoluminal treatments.	• Severity, stage, or site, or type	Not stated explicitly	• Patients with advanced NSCLC

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Dong et al., 2007[16]	Objectives: To determine the effectiveness and safety of elemene in the treatment of patients with lung cancer.	No	NA	Authors planned to study: • Stage of disease, for example Stage III and IV Inclusion Criteria: • Patients with histologically proven lung cancer (small cell or non-small cell), at any stage of their illness • Lung cancer patients who received elemene therapy alone or combined with other conventional therapies such as chemotherapy, radiotherapy, surgery, physical therapy, TCM at the same time, regardless of mode of delivery (oral, injection, infusion)
CANCER, PROSTATE				
Shelley et al., 2007[17]	Objective: To compare the efficacy and side effects of cyrotherapy with other primary treatments in the management of patients with localised prostate cancer	No	NA	• Men with localized prostate cancer Looked at in analysis: • Age • Stage of prostate cancer • Pre-op PSA • Gleason score

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
De Conti et al., 2007[18]	Objective: Evaluate the effectiveness and safety of IAS compared to continuous androgen suppression for treating prostatic cancer	No	NA	Inclusion criteria: • Patients diagnosed with prostate cancer who have not received prior androgen suppression therapy Participants were grouped by: • Early primary therapy for clinically localized disease • Clinically advanced disease and no prior therapy • Adjuvant therapy in high-risk patients with clinically localized disease and treated with either RP or RT • PSA or clinical evidence of failure following definitive therapy

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HEART FAILURE				
Frobel et al., 2009[19]	Objectives: To assess the effect of beta-adrenoceptor-blockers in children with congestive heart failure.	• Age	Not stated explicitly	• Age (infants and toddlers age 28 days to 23 months; and children up to 18 years of age) • Aetiology of heart failure • Severity of heart failure • Additional organ diseases
CESAREAN SECTION				
Mathai and Hofmeyr[20]	Objective: To determine the benefits and risks of alternate methods of abdominal surgical incisions for cesarean section	No	NA	• Inclusion criteria: • Pregnant women due for delivery by cesarean section. Planned subgroup analyses: • Primary, repeat and mixed or undefined cesarean sections • General, regional and mixed or undefined anaesthesia
Anorlu Rose and Hofmeyr, 2008[21]	Objective: To compare the effects of manual removal of the placenta with cord traction at cesarean section	No	NA	Inclusion criteria: • Women undergoing a cesarean, whether emergency or elective

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CHRONIC KIDNEY DISEASE				
Navaneethan et al., 2009[22]	Objective: To evaluate the benefits and harms of statins in CKD patients who were not receiving renal replacement therapy.	• Co-interventions	Not stated explicitly	• Age • Sex • Baseline renal pathology and degree of renal impairment • Presence of cardiovascular comorbidities • Restricted to pre-dialysis CKD patients
Roderick et al., 2007[23]	Objectives: O1: Does correction of metabolic acidosis improve the nutritional state of CKD patients? O2: Does the correction of metabolic acidosis alter bone turnover and so reduce the development of renal osteodystrophy in CKD patients? O3: Is the use of oral bicarbonate to correct metabolic acidosis safe in relation to hypertension and fluid overload? O4: Does the correction of metabolic acidosis improve patients' quality of life, reduce hospitalisation or reduce mortality?	No	NA	Selection criteria: • Adults or children with CKD, whether or not they were receiving RRT • Presence of metabolic acidosis at entry to trial (initial venous bicarbonate must be stated). Addressed in analysis: • Age • Sex • Comorbidities (diabetes) • Ethnicity • Time on dialysis • Duration of dialysis

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
COPD				
Effing et al., 2007[24]	Objectives: I. To evaluate whether self-management education programmes in COPD lead to improved health outcomes II. To evaluate whether self-management education programmes in COPD lead to a reduction in health care utilisation	No	NA	• Age • Sex • Socio-economic status • Disease severity Participant criteria: • Patients with a clinical diagnosis of COPD and not asthma as primary diagnosis were included.
Puhan al., 2009[25]	Objective: To assess the effects of pulmonary rehabilitation after COPD exacerbations on future hospital admissions (primary outcome) and other patient-important outcomes (mortality, health related quality of life and exercise capacity)	No	NA	• Age • Sex • FEV1 score Patient criteria: • COPD patients after in- or out-patient care for acute exacerbation. More than 90% of study participants were required to be COPD patients.

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Puhan et al., 2007[26]	Objective: Analyse randomized controlled trials investigating the clinical benefit of antibiotics for COPD exacerbations	No	NA	Selection criteria: • Patients suffering from an acute exacerbation defined as a worsening of a previous stable situation with symptoms such as increased dyspnea, increased cough, increased sputum volume or change in sputum colour. • Studies must have >90% of patients had a clinical (physician-based) diagnosis of COPD, or. Ideally, spirometrically confirmed COPD. • Did not look at subgroups in analysis.

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Yang et al., 2007[27]	Objective: To determine the efficacy of regular use of inhaled corticosteroids in patients with stable COPD	• Severity, stage, or site	Not stated explicitly	Considered in analysis: • Age • Sex • Ethnicity • Smoking history • Severity of COPD Participant criteria: • Adults with COPD defined as progressive chronic airflow limitation without recent exacerbation, hospitalization, or need for antibiotics or systemic steroids

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
DEPRESSION				
Nieuwenhuijsen et al., 2008[28]	Objective: To evaluate the effectiveness of interventions aimed at reducing work disability in depressed workers	No	NA	Patient inclusion characteristics: • Adults (over 17 years) • Workers (employed or self-employed) • Diagnosed for dysthymic disorder, minor depressive disorder or MDD according to the DSM-IV, RDC, or ICD-10 Included in analyses: • Age • Sex • Marital status (single, married, cohabitating, divorced or separated) • Age at onset of depression • Duration of depression • Previous treatment • Alcohol/drug dependence • Ability or inability to read or write • Employed full or part-time

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Morriss et al., 2007[29]	Objectives: • To compare the effectiveness of an EWS intervention plus TAU vs. TAU not involving a psychological therapy on time to manic, depressive and all bipolar episodes. • To compare the effectiveness of an EWS intervention plus TAU vs. TAU plus another psychological therapy on time to manic, depressive and all bipolar episodes. • To compare the effectiveness of intermittent medication used on recognition of EWS without continued mood stabilizing medication verses TAU involving continued mood stabilizing medication on time to manic, depressive and all bipolar episodes.	No	NA	Inclusion criteria: • Adults with a diagnosis of bipolar disorder or associated diagnoses based on standardized psychiatric criteria (RDC, DSM-IV or ICD-10) Subgroups in analyses: • Age • Sex • Number of previous episodes

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Hackett et al., 2008[30]	Objective: To determine if pharmaceutical or psychological interventions can prevent depression and improve physical and psychological outcomes in patients with stroke	No	NA	Participant inclusion criteria: • Participants with a confirmed history of stroke, defined according to clinical criteria to include cerebral infarction, intracerebral haemorrhage, and uncertain pathological subtypes. Subgroup analyses: • Demographics (age, sex) • Stroke severity • Stroke sequence (1st vs. recurrent) • Time of stroke onset • Prior history of psychiatric illness • Current neurological status • History of coronary artery disease
Furtado et al., 2008[31]	Objective: To assess the effects of atypical antipsychotic drugs on people who have a diagnosis of both schizophrenia and depression	No	NA	Selection criteria: • People diagnosed with both schizophrenia and depression Subgroups in analyses: • Age • Sex

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Cipriani et al., 2009[32]	Objectives: 1. The primary objective of the review was to determine the effectiveness of olanzapine compared with placebo or other active treatment, either as monotherapy or as adjunctive treatment to lithium, valproate or adjuctive compounds in: (a) preventing manic, depressive and mixed episodes of bipolar affective disorder (b) preventing episodes in patients with rapid cycling disorder bipolar disorder 2. To review the effect of olanzapine on patients' general health and social functioning 3. To review the acceptability of long-term olanzapine treatment to patients, measured by numbers and reasons for withdrawal from treatment, by adherence and by reference to patients' expressed views regarding treatment 4. To investigate the adverse effects of olanzapine, including general prevalence of adverse events 5. To determine overall mortality rates on long-term treatment with olanzapine	No	NA	Patient inclusion criteria: • Patients diagnosed with bipolar disorder. Subgroups in analyses: • Age • Age at onset • Diagnosis (bipolar disorder I, bipolar disorder II or schizoaffective disorder bipolar type)

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HEAVY MENSTRUAL BLEEDING				
Liu et al., 2009[33]	Objectives: Primary objective: To assess the benefits and risks of herbal preparations for treating uterine fibroids. Secondary objective: To assess participant compliance in the use of herbal preparations for treating uterine fibroids.	No	NA	• Age • Ethnicity • Symptoms • Parity • Level of education • Duration of diagnosis • BMI • Hemoglobin level • Number of fibroids Participant inclusion criteria: • Women with uterine fibroids diagnosed by clinical symptoms and physical signs, and confirmed by ultrasound scanning, CT, MRI, or a combination of more than one of the procedures.

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Beaumont et al., 2007[34]	Objective: To determine the effectiveness and tolerability of Danazol when used for heavy menstrual bleeding in women in reproductive years. To investigate- • Whether treatment with Danazol is more effective than placebo in reducing heavy MBL. • Whether treatment with Danazol is more effective than other medical therapies (antifibrinolytics, NSAIDs, progestogens) in reducing heavy MBL. • If effective, what is the optimum dosage of Danazol. • Whether treatment with Danazol leads to an improved quality of life for women with heavy MBL. • Whether women tolerate treatment with Danazol and find it an acceptable treatment.	No	NA	Selection criteria: • Women of reproductive years • Regular (21-35 days cycle) heavy MBL, subjectively or objectively defined (for example by alkaline haematin method) Subgroup analyses: • Age • MBL in ml/cycle

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Lethaby et al., 2007[35]	Objectives: To determine the effectiveness, safety and tolerability of NSAIDs in achieving a reduction in MBL in women of reproductive years with heavy menstrual bleeding. Tested the following hypotheses • Treatment with NSAIDs is more effective than placebo in reducing MBL • Treatment with NSAIDs is more effective than other medical therapies (anti-fibrinolytics, danazol, hormone treatment, LHRH/GNRH analogues) in reducing MBL. • Individual NSAIDs have similar efficacy in reducing MBL. • Treatment with NSAIDs is associated with a lower incidence of adverse events and higher adherence and greater acceptability than other medical therapies. • Treatment with NSAIDs is a cost-effective method of treating heavy menstrual bleeding. • Treatment with NSAIDs leads to an improved quality of life for women with heavy menstrual bleeding (in particular, an improvement in symptoms of dysmenorrhoea).	No	NA	• Age • Diagnosis • MBL in ml/cycle Selection criteria: • Women of reproductive years with regular heavy periods measured either objectively or subjectively and no pathological or iatrogenic (treatment induced) causes for their heavy MBL

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Lethaby et al., 2008[36]	Objective: To determine the effectiveness of oral progestogen therapy in achieving a reduction in MBL in women of reproductive age with heavy menstrual bleeding. Tested the following hypotheses • Treatment with oral progestogens is more effective than placebo in reducing MBL. • Treatment with oral progestogens is more effective than other medical therapies (antifibrinolytics, danazol, combined oral contraceptives, progesterone and the progesterone-releasing IUS, NSAIDs, GNRH analogues) in reducing MBL. • Treatment with oral progestogens is associated with a lower incidence of adverse events and higher compliance and acceptability than with other medical therapies. • Treatment with oral progestogens is more cost effective than other medical therapies in treating menorrhagia. • Treatment with oral progestogens leads to an improved quality of life for women with menorrhagia.	No	NA	Selection criteria: • Women of reproductive age with regular heavy periods measured either objectively or subjectively and no pathological or iatrogenic (treatment induced) causes for their heavy MBL Subgroups in analysis: • Age • MBL in ml/cycle

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HYPERTENSION				
Wiysonge et al., 2007[37]	Objectives: • To quantify the effects of beta-andrenergic blocking agents used as first-line treatment or monotherapy on morbidity and mortality in adults with hypertension. • To determine whether the effects on morbidity and mortality are similar to those of other classes of anti-hypertensive drugs. • To determine whether the use of first-line beta-blocker therapy is associated with an increased incidence of adverse effects when compared to placebo or other classes of antihypertensives. • To determine whether the effects of first-line beta-blockade differ by type and dose of beta-blocker (e.g., cardioselective vs. non-selective beta-blocker), and by age or "ethnicity" (e.g., blacks vs. whites) of patients.	• Age, race/ethnicity	Not stated explicitly	• Age • Ethnicity (e.g., black vs. white) • Comorbid conditions • Baseline blood pressure Types of participants: • Men and non-pregnant women, aged 18 years and over, with hypertension as defined by cut-off points operating at the time of the study under consideration.
Hodson et al., 2007[38]	Objectives: To assess the benefits and harms of different corticosteroid regimens in children with SSNS. The benefits and harms of therapy were studied in two groups of children: • Children in their initial episode of SSNS • Children who experience a relapsing course of SSNS	• Age • Severity, stage, or site	Not stated explicitly	• Stage of SSNS (initial vs. relapse) • Age Selection criteria: • Children (3 months to 18 years) in their initial or subsequent episode of SSNS

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Abalos et al., 2007[39]	Objective: To determine the possible benefits, risks and side-effects of anti-hypertensive drug treatments for women with mild to moderate hypertension during pregnancy (defined whenever possible as a systolic blood pressure of 140 to 169 mmHg or diastolic blood pressure of 90 to 109 mmHg, or both). Also, to compare the differential effects of alternative drug regimens. The comparisons of: • any antihypertensive drug with either no drug or placebo; • one antihypertensive drug compared with another. For this review, the commonly used drugs are regarded as control and compared with other agents (for example, any antihypertensive vs. methyldopa, any antihypertensive vs. calcium channel blockers).	• Sex • Severity, stage, or site • Pregnancy	Not stated explicitly	Selection criteria: • Women with mild to moderate hypertension during pregnancy, defined, whenever possible, as systolic blood pressure 140 to 169 mmHg and diastolic blood pressure 90 to 109 mmHg. Subgroups analyzed: • Type of hypertensive disorder (mild to moderate alone, mild to moderate with proteinuria, chronic hypertension, unspecified) • Gestational age (less than 32 weeks' gestation, about 32 weeks or more gestation, or unclassified/mixed)

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Zijdenbos et al., 2009[40]	Objectives: To evaluate the efficacy of psychological interventions for the treatment of IBS.	No	NA	Subgroups in analysis: • Age • Gender • Duration of complaints • Presence of DSM diagnosis • Criteria for diagnosis (Rome, Rome I, Clinical or Manning) • Subtype of IBS • Severity of IBS • Previous treatment Inclusion criteria: • Patients over 16 years of age
Webb et al., 2007[41]	Objective: To evaluate the efficacy of hypnotherapy for the treatment of IBS.	No	NA	Participant inclusion criteria: • Patients of any gender, age or ethnic origin with a diagnosis of IBS and who did not have an organic cause for their gastrointestinal symptoms. Subgroups looked at in analysis: • Gender • Age • Comorbidities • Duration of IBS

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Evans et al., 2007[42]	Objective: To evaluate the efficacy and tolerability of tegaserod for the treatment of IBS and chronic constipation in adults and adolescents aged 12 years and above.	• Age	Not stated explicitly	Inclusion criteria: • Adults and adolescents aged 12 years or above with diagnosis of: - IBS according to any predefined/specified diagnostic criteria (e.g., Manning, Rome [I,II,III]); or - Chronic constipation Subgroup analysis: • Gender • Age • Duration of symptoms • Duration of IBS

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
LABOR INDUCTION				
Smyth et al., 2007[43]	Objectives: To determine the effectiveness and safety of amniotomy alone for (1) routinely shortening all labours that start spontaneously, and (2) shortening labours that have started spontaneously, but have become prolonged.	• Severity, stage, or site	Not stated explicitly	• Parity: primigravid women compared with parous women; • Previous mode of delivery: cesarean section compared with vaginal delivery and no previous delivery; • Stage of labour: less than 3 cm dilated at time of amniotomy compared with 3 cm or more; • Fetal surveillance: continuous fetal heart monitoring compared with intermittent; • Pain relief: pharmacological compared with non-pharmacological • Indication for intervention: dysfunctional labour vs. routine use or fetal compromise • Position in labour: mobile vs. restricted movement in women without an epidural

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Smyth et al., 2007[43] (continued)				Inclusion criteria: • Pregnant women with singleton pregnancies regardless of parity and gestation at trial entry in spontaneous labour.
Boulvain et al., 2008[44]	Objectives: To determine the effects of intracervical prostaglandins for third trimester cervical ripening or induction of labour compared with placebo/no treatment and with vaginal prostaglandins (except misoprostol).	No	NA	Predefined group analyses: • Previous cesarean section or not; • Nulliparity or multiparity • Membranes intact or ruptured • Cervix unfavourable, favourable or undefined Inclusion criteria: • Pregnant women due for third trimester induction of labour, carrying a viable fetus.

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
MYOCARDIAL INFARCTION				
Li, et al., 2007[45]	Objectives: In patients suffering acute myocardial infarction: • To examine the effect of intravenous magnesium vs. control on early mortality (primary objective), stratified by time since onset of symptoms (<6 hours, 6+ hours), use of thrombolysis (used, not used), dose of magnesium used (<75 mmol, 75+ mmol) • To examine the effect of intravenous magnesium vs. control on early morbidity (secondary objective), including ventricular fibrillation and tachycardia, atrioventricular block, bradycardia, heart failure, cardiogenic shock, hypotension, severe arrhythmia needing treatment or Lown 2-5, and flushing	No	NA	Participant inclusion criteria: • All patients with first-time acute myocardial infarction or suspected myocardial infarction in the preceding 24 hours diagnosed by clinical symptoms, enzymes and ECG, regardless of age, gender, infarct size and location, and without contraindication to magnesium. Pre-determined data abstraction: • Age • Gender • Comorbid conditions

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Magee et al., 2008[46]	Objective: To determine the effect of heparins (UFH and LMWH) compared with placebo for the treatment of patients with ACS.	No	NA	Inclusion criteria: • Adult patients (> 18 years of age) presenting with acute coronary syndromes requiring treatment within 72 hours of presentation of their last episode of chest pain. • Age • Gender • Time to presentation
Wu et al., 2008[47]	Objective: To assess the effects (both benefits and harms) of danshen preparations with AMI.	No	NA	Participant inclusion criteria: • Men and women of any age or ethnic origin with AMI defined as the presence of unequivocal ECG changes and/or unequivocal enzyme changes. Subgroups in analysis: • Age • Gender

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
OSTEOARTHRITIS				
Brouwer et al., 2007[48]	Objective: To assess the effectiveness and safety of an osteotomy for treating osteoarthritis of the knee.	• Severity, stage, or site	Not stated explicitly	Selection criteria: • Adult patients (>18 years) with unicompartmental osteoarthritis of the medial or lateral compartment of the knee confirmed by radiographic or arthroscopic investigation. Subgroups in analysis: • Age
Fransen et al., 2008[49]	To determine whether land-based therapeutic exercise is beneficial for people with knee OA in terms of reduced joint pain or improved physical function.	• Severity, stage, or site	Not stated explicitly	• Sex • Age • ACR criteria • Years spent sedentary • Lequensne score Participant inclusion criteria: • Adults with either an established diagnosis of knee OA according to accepted criteria or self-reporting knee OA on the basis of chronic joint pain (without radiographic confirmation).

Evidence Table C3. Cochrane final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Laupattarakasem et al., 2008[50]	Main objective- To estimate the effectiveness of AD on knee OA pain reduction (reduced use of relevant medications) and/or functional improvement Secondary objectives- • The type or stage of severity of the OA in which AD is most effective. • The expected length of effectiveness until the patients need further intervention.	Severity, stage, or site	Not stated explicitly.	• Demographics (age, sex, ethnic group) • Side/location of lesions in knee • Diagnostic criteria (as defined by American College of Rheumatology vs. not stated) • Staging criteria of the OA • Duration of disease • Co-morbidities Patient inclusion criteria: • Patients with a diagnosis of primary or secondary OA of the knees, who did not have other joint involvement or conditions requiring long term use of NSAIDs.

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, BREAST				
Boutin et al., 2007[51]	Objective: Examine the extent to which CBT, SEGT, and a combination of these two treatments impact women with breast cancer.	No	NA	• Participant inclusion: • Women whose current diagnosis represented all stages of breast cancer. Subgroups analyzed: • Cancer type (metastatic, newly diagnosed or Stage 0, I or II)
van der Ploeg et al., 2007[52]	Objective: The purpose of the present review is to compare the results of recent studies investigating RFA for the treatment of breast cancer.	No	NA	• Age • Diagnosis (CNB or FNA) • Tumour characteristics • Site of tumour
CANCER, LUNG				
Yau et al., 2007[53]	Research questions: Should LDCT be introduced for screening of lung cancer in a high-risk population? Three principle outcomes were assessed: (1) the operating characteristics of LDCT for screening of lung cancer, (2) the percentage of LDCT detected lung cancers at an early stage (Stage I), and (3) the potential reduction in lung cancer mortality	• Risk factors • Severity, stage, or site	Not stated explicitly	• Age • Gender • Smoking history Study/participant criteria: • Former and current smokers

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Micames et al., 2007[54]	Objective: To estimate the diagnostic accuracy of EUS-FNA for staging mediastinal lymph nodes (N2/N3 disease) in patients with lung cancer.	• Severity, stage, or site	Not stated explicitly	Participant criteria: • Adult patients (> 18 years old) with suspected or previously diagnosed NSCLC for staging of mediastinal lymph nodes
Coory et al., 2008[55]	Objective: To evaluate and critically appraise the effectiveness of multidisciplinary teams for lung cancer.	No	NA	• Age • Gender • Stage of disease
CANCER, PROSTATE				
Tarnhuvud et al., 2007[56]	Objective: To investigate what nurses do to improve the health of men who are receiving radiotherapy treatment due to prostate cancer	No	NA	• NA
HEART FAILURE				
Lago et al., 2007[57]	Objective: To examine the risk of congestive heart failure and of cardiac death in patients given TZDs.	No	NA	• Age • Sex • BMI • Baseline HbA2c • Baseline medical history (HTN, HLD, CAD, CHF, CKD or nephropathy) Inclusion criteria: • Patients given TZDs.
Nasr et al., 2007[58]	Objective: To estimate the preventive efficacy of beta blocker treatment on AF occurrence in patients with heart failure.	No	NA	• Age • Aeitiology of HF • LVEF(%) Inclusion Criteria: • Patients with CHF

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Roberts et al., 2007[59]	KQ: Is metformin safe to use in patients with heart failure?	No	NA	• Age • Gender • Race • Left ventricular function • Cardiac history • Diabetes complications • History of cerebrovascular accident • Discharge medication • Chronic lung disease • Dementia • Glucose value in mg/dL Inclusion criteria: • Patients with heart failure
CESAREAN SECTION				
Nicholson et al., 2008[60]	KQ1: What is the evidence for the risks and benefits of oral diabetes agents (e.g., second-generation sulfonylureas and metformin), as compared to all types of insulin, for both the mother and neonate in the treatment of women with gestational diabetes? a. How does maternal outcome vary based on the level of glucose at the initiation of a medication? b. How does neonatal outcome vary based on the level of glucose at the initiation of a medication?	• Risk factors	Not stated explicitly	Inclusion criteria: • Women with gestational diabetes confirmed by an OGTT • Waist circumference and BMI • Gestational age at diagnosis of gestational diabetes • Progesterone-only contraception use • Prior history of gestational diabetes • Metabolic risk factors • Lifestyle factors (e.g., physical activity) • Age

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: What is the evidence that elective cesarean delivery or the choice of timing of induction in women with gestational diabetes results in beneficial or harmful maternal and neonatal outcomes? a. What is the evidence for elective cesarean delivery at term, as compared to an attempt at vaginal delivery (spontaneous or induced) at term, with regard to beneficial or harmful maternal and neonatal outcomes in gestational diabetes? i. cesarean vs. spontaneous labor and vaginal delivery ii. cesarean vs. induced labor and vaginal delivery iii. cesarean vs. any attempt at vaginal delivery at term b. What is the evidence for labor induction at 40 weeks, as compared to labor induction at an earlier gestational age (less than 40 weeks) or spontaneous labor, with regard to beneficial or harmful maternal and neonatal outcomes in gestational diabetes? i. labor induction at less than 40 weeks vs. labor induction at 40 weeks ii. labor induction at 40 weeks vs. spontaneous labor iii. labor induction at less than 40 weeks vs. spontaneous labor	• Risk factors	Not stated explicitly	Inclusion criteria: • Women with gestational diabetes confirmed by an OGTT

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	c. How is the EFW related to outcomes of management of gestational diabetes with elective cesarean delivery of the timing (e.g., gestational age range) of labor induction? d. How is gestational age related to outcomes of management of gestational diabetes with elective cesarean delivery or the choice of timing (i.e., gestational age range) of labor induction?			
	KQ3: What risk factors, including but not limited to family history, physical activity, pre-pregnancy weight, and gestational weight gain, are associated with short-term and long-term development of type 2 diabetes following a pregnancy with gestational diabetes?	• Risk factors	Not stated explicitly	• Risk factors considered independent variable not subgroup analyses. Inclusion criteria: • Women with gestational diabetes confirmed by an OGTT
	KQ4: What are the performance characteristics (sensitivity, specificity, and reproducibility) of tests for diagnosing type 2 diabetes after pregnancy in patients with a history of gestational diabetes? Are there differences in the performance characteristics of the test results based on subgroup analysis?	• Risk factors	Not stated explicitly	• Family history • Insulin required during pregnancy Inclusion criteria: • Women with gestational diabetes confirmed by an OGTT
Press et al., 2007[61]	Objective: To compare the prevalence of postpartum urinary incontinence after cesarean section compared with vaginal birth	No	NA	NA

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CHRONIC KIDNEY DISEASE				
Balamuthusamy et al., 2008[62]	Objective: Analyze the effects of RAS blockade on CV outcomes in patients with CKD	No	NA	Inclusion criteria: • Patients with CKD defined as CKD stage 2 and above per the KDOQI guidelines. • Etiology of CKD (all nephropathy, proteinuria, diabetic nephropathy, nondiabetic nephropathy, hypertensive nephropathy)

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Strippoli et al., 2008[63]	Objective: To analyze the benefits and harms of statins in patients with chronic kidney disease (pre-dialysis, dialysis, and transplant populations).	• Severity, stage, or site	Not stated explicitly	• Age • Sex • Race • Stage of chronic kidney disease (pre-dialysis, dialysis, and transplant) • Baseline risk covariates Inclusion criteria: • Patients with CKD who were having maintenance dialysis treatment, had had renal transplantation, had an elevated baseline mean serum creatinine (>1.4 mg/dl (0.121 mmol/l) or as defined by authors), of had an impairment of the glumerular filtration rate as defined by the kidney disease outcome quality initiative guidelines with values of glomerular filtration rate <60 ml/min/1.73 m^2 (stages 3-5) or >60 ml/min/1.73 m^2 along with other markers of kidney damage such as proteinuria.

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
COPD				
Bradley-Drummond et al., 2008[64]	Objective: To systematically review and quantitatively synthesize the effects of ICS therapy on mortality and adverse events in patients with stable COPD	• Severity, stage, or site	Not stated explicitly	Inclusion criteria: • Adults (≥40 years) with COPD (defined as clinical diagnosis of COPD or as current or former smoking [≥10 pack years] and an FEV_1 to forced vital capacity ratio <0.70) Subgroup analyses: • Baseline COPD severity
Niesink et al., 2007[65]	Objective: To investigate effectiveness of chronic disease management programmes on the quality-of-life of people with COPD	No	NA	Inclusion criteria: • Outpatients with COPD • Sex • Age • Degree of airflow limitation measured by FEV_1
Quon et al., 2008[66]	Objective: Systematically evaluate the effectiveness of systemic corticosteroids, antibiotics, and NPPV for patients with acute exacerbation of COPD.	• Severity, stage, or site	Not stated explicitly	Inclusion criteria: • Adults (>19 years of age) with acute COPD exacerbations • Age • FEV_1 level • Baseline arterial blood gas measurement
Singh et al., 2008[67]	Objective: To ascertain the cardiovascular risk of inhaled anticholinergics, including cardiovascular death, MI, and stroke	No	NA	• Age • Sex • Current, former or non-smoker • Severity of COPD • Preexisting cardiac disease or cardiovascular risk factors

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Sobieraj et al., 2008[68]	Objective: To elucidate the benefits and risks associated with adjunctive ICS treatment in patients with severe or very severe COPD	• Severity, stage, or site	Not stated explicitly	Inclusion criteria: • Patients with severe or very severe COPD and a forced expiratory volume in 1 second (FEV_1) <80% and an FEV_1/forced vital capacity ratio <70% • Sex • Age • Smoking status • Number of pack years • Severity of COPD
DEPRESSION				
Usala et al., 2008[69]	Objective: To evaluate the efficacy of SSRIs in children and adolescents with depressive disorder	• Age	Not stated explicitly	Inclusion criteria: • Children and adolescents with depressive disorder or depressive symptoms • Age • Comorbidities • Severity of depression
Cuijpers et al., 2008[70]	Objective: Examine the effects of psychological treatments on PPD compared to control conditions and to other (nonpsychological) interventions	No	NA	Inclusion criteria: Adult female participants with postpartum depression as diagnosed through clinical interview and/or self-report questionnaire.
Chin, 2007[71]	Objective: To examine the clinical effects of reminiscence therapy on the life satisfaction, happiness, depression and self-esteem of older adults aged 50 or above	• Age	Not stated explicitly	Participant criteria: • Older adults of age 50 years or above. Subgroup analysis: • Age was reported • Authors noted in the "methodological issues" section that the limited number of included studies prohibited subgroup analysis.

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Barbui et al., 2008[72]	Objective: To determine the effectiveness and acceptability of paroxetine.	No	NA	Participant criteria: • Adults (≥18 years of age) of either sex with a diagnosis of major depression according to any diagnostic criteria. • Severity of depression
DYSPEPSIA				
Hiyama et al., 2007[73]	Objective: To perform a meta-analysis of the effects of prokinetic agents in patients with functional dyspepsia.	No	NA	• Age • Gender Inclusion criteria: • Patients with functional dyspepsia.
Jin and Lim, 2007[74]	Objective: To evaluate the effect of eradicating Heliobacter pylori on dyspeptic symptoms in patients with functional dyspepsia in China.	No	NA	Inclusion criteria: • Patients with functional dyspepsia.
Wang et al., 2007[75]	Objective: To assess systematically the efficacy of PPIs in the treatment of functional dyspepsia compared with placebo and to determine if any difference in the response exists between symptom subgroups of functional dyspepsia.	• Severity, stage, or site	Not stated explicitly	• Age • Sex • History of dyspepsia • Dyspepsia symptoms (i.e., ulcer-like, reflux-like, dysmotility-like, and un-specified dyspepsia) • H. pylori status Inclusion criteria: • Patients with functional dyspepsia defined as persistent or recurrent dyspepsia with no evidence of organic disease to explain patient's symptoms

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HYPERTENSION				
Horvath et al., 2008[76]	Review question: To assess the long-term effects of (1) dietary interventions intended to reduce body weight, (2) pharmacologically induced reduction in body weight, and (3) reduction of body weight through invasive interventions on all causes of death, cardiovascular morbidity, adverse events, and BP in people with essential hypertension	No	NA	No. Planned to perform subgroup analyses for the following items. However because of low number of included trials and lack of information, subgroup analyses were not possible. • Sex • Age • BMI • Concomitant diseases • Ethnicity • BP at baseline • BP goals • Concomitant antihypertensive therapy • Socioeconomic status Inclusion criteria: • Patients with essential hypertension aged 18 years or older (excluding pregnant women).
Connell et al., 2008[77]	KQ1: Are community interventions effective in reducing hypertension or increasing knowledge about hypertension among black groups of African descent? KQ2: Are there components of community interventions that demonstrate greater effectiveness than others? KQ3: Do features aimed at achieving cultural sensitivity in interventions increase the effectiveness of those interventions?	• Race	Based on statistics in the US and UK, people of African descent have higher stroke incidence and higher hypertension prevalence then other groups.	Inclusion criteria: • Black adults (18 years and over) of African descent. • Studies aimed at hypertension control needed to include participants with hypertension (BP > 140/90 mmHg), but not studies aimed at improving hypertension knowledge.

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
IBS				
Ford et al., 2008[78]	Objective: To systematically review the literature of the accuracy of individual symptoms and combinations of findings in diagnosing IBS.	No	NA	Inclusion criteria: • adults (16+ years old) with lower gastrointestinal tract symptoms who were attending for investigation (colonoscopy, barium enema, or computed tomographic colography) Authors note that several of the studies reviewed included prevalence by: • age • gender
Ford et al., 2008[79]	Objective: To determine the effect of fibre, antispasmodics, and peppermint oil in the treatment of IBS.	No	NA	Inclusion criteria: • Adults (>16 years) with a diagnosis of IBS based on clinician's opinion or that met specific diagnostic criteria (Manning, Kruis score, Rome I, II, or III), combined with the results of investigations to exclude organic disease if trial investigators thought this necessary. • Gender • Subtype of IBS

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Ford et al., 2009[80]	Objective: To estimate prevalence of celiac disease in unselected adults who met diagnostic criteria for IBS.	• Risk factors	Not stated explicitly	Inclusion criteria: • Adults (90% aged ≥ 16 years) with a presumed diagnosis of IBS according to physician opinion, questionnaire findings, or normal findings at examination who met specific diagnostic criteria such as those of Manning et al, Rome I, II, or III criteria, or the scoring system of Kruis et al.
Rahimi et al., 2008[81]	Objective: To evaluate the efficacy of SSRIs for the management of IBS	No	NA	• Age • Sex • IBS subtype
MYOCARDIAL INFARCTION				
Baker and Couch, 2007[82]	KQ1: What is the effect of the macrolide antimicrobial, azithromycin, on clinical outcomes in patients with CAD?	No	NA	Inclusion criteria: • Patients with secondary CAD
Ioannidis and Katritsis, 2007[83]	KQ1: Compare PCI with medical therapy in stable patients with an occluded artery 1 to 45 days after MI.	• Severity, stage, or site	Not stated explicitly	Inclusion criteria: • Stable patients with previous MI with angiographic evidence of persistent occlusion in the culprit vessel. • Age • Gender
Sinno et al., 2007[84]	KQ1: Compare the safety and efficacy of adjunctive use of reduced-dose thrombolytics and Gp IIbII inhibitors to the sole use of Gp IIbIIa inhibitors before PCI in patients presenting with acute STEMI.	No	NA	Inclusion criteria: • Patients presenting with acute STEMI

Evidence Table C4. Database of Abstracts of Reviews of Effects (DARE) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
OSTEOARTHRITIS				
Christensen et al., 2008[85]	Objective: To evaluate the efficacy of preparations with ASUs in OA patients using meta-analysis on RCTs	No	NA	• Site of OA (hip or knee) • Duration of trial • Age • Sex • BMI • KL score • Pain • Lequesne index (measures severity of osteoarthritis) Inclusion criteria: • Patients with clinical or radiographic evidence of OA.
Minns et al., 2007[86]	Objective: To evaluate the effectiveness of physiotherapy exercise after elective primary total knee arthroplasty in patients with osteoarthritis.	• Severity, stage, or site	Not stated explicitly	• Included patients discharged from hospital after elective primary total knee arthroplasty for osteoarthritis. Subgroup analysis: • Time since surgery (3 to 4 months vs. 12 months)
Evidence Pisters et al., 2007[87]	Objective: To determine the long-term effectiveness ($\geq$ 6 months after treatment) of exercise therapy on pain, physical function, and patient global assessment of effectiveness in patients with OA of the hip and/or knee.	• Severity, stage, or site	Not stated explicitly	NA

Table C5. Drug Effectiveness Review Program (DERP) reports

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
HEART FAILURE	NOTE: The DERP heart failure reports are two separate reports with different authors. On the DERP website, one report falls under "ACE Inhibitors" and the other under "Angiotensin II Receptor Antagonists" Both reports use the same key questions, and both had their key questions updated in 2009. Thus, we are using the most recent (and identical) questions.			
Chou et al., 2005[88]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease or diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with ACE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults 18+ years)
	KQ2: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in effectiveness and efficacy between DRI, ACE-I and AIIRA drugs? 2a. When used as monotherapy? 2b.When used in combination with one another?	• Co-interventions	Not stated explicitly	• Age (adults 18+ years)
	KQ3: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in harms between DRI, ACE-I, and AIIRA drugs?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ4: Are there subgroups based on demographics (age, racial groups, gender), other medications, or co-morbidities for which there are inter-class differences between DRI, ACE-I and AIIRA drugs?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities
Furmaga et al., 2006[89]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with ACE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults 18+ years)
	KQ2: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in effectiveness and efficacy between DRI, ACE-I and AIIRA drugs? 2a. When used as monotherapy? 2b. When used in combination with one another?	• Co-interventions	Not stated explicitly	• Age (adults 18+ years)
	KQ3: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in harms between DRI, ACE-I, and AIIRA drugs?	No	NA	NA
	KQ4: Are there subgroups based on demographics (age, racial groups, gender), other medications, or co-morbidities for which there are inter-class differences between DRI, ACE-I and AIIRA drugs?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities
Helfand et al., 2009[90]	KQ1: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine or bleeding esophageal varices, do beta blocker drugs differ in effectiveness/efficacy?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine prophylaxis or bleeding esophageal varices, do beta blocker drugs differ in harms?	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, gender), other medications (drug-drug interactions), or co-morbidities (drug-disease interactions) for which one beta blocker is more effective or associated with fewer adverse effects?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities
CHRONIC OBSTRUCTIVE PULMONARY DISEASE				
Jonas et al., 2008[91]	KQ1: What is the comparative efficacy and effectiveness of controller medications used to treat outpatients with persistent asthma?	No	NA	NA
	KQ2: What is the comparative tolerability and frequency of adverse events for controller medications used to treat outpatients with persistent asthma?	No	NA	NA
	KQ3: Are there subgroups of these patients based on demographics (age, racial groups, gender), asthma severity, comorbidities (drug-disease interactions, including obesity), smoking status, genetics, or pregnancy for which asthma controller medications differ in efficacy, effectiveness, or frequency of adverse events?	• Risk factors • Age, race, sex • Severity, stage, or site • Co-interventions • Comorbidities • Pregnancy	Not stated explicitly	• Demographics (age, racial groups, gender) • Asthma severity • Comorbidities (drug-disease interactions, including obesity) • Smoking status • Genetics • Pregnancy
Hansen et al., 2006[92]	KQ1: For outpatients with asthma or COPD, do inhaled corticosteroids differ in effectiveness?	No	NA	NA
	KQ2: For outpatients with asthma or COPD, do inhaled corticosteroids differ in safety or adverse events?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, and sex), other medications, comorbidities, or pregnancy for which one inhaled corticosteroid is more effective or associated with fewer adverse events than another?	• Risk factors • Age, race, sex • Severity, stage, or site • Co-interventions • Comorbidities • Pregnancy	Not stated explicitly	• Demographics (age, racial groups, and sex) • Other medications • Comorbidities • Pregnancy
DEPRESSION				
Gartlehner et al., 2008[93]	KQ1: For outpatients with depressive, anxiety, and/or premenstrual dysphoric disorders, do second-generation antidepressants differ in efficacy or effectiveness?	No	NA	NA
	KQ2: For outpatients with depressive, anxiety, and/or premenstrual dysphoric disorders, do second-generation antidepressants differ in safety or adverse events?	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, and sex), other medications, or comorbidities for which one second-generation antidepressant is more effective or associated with fewer adverse events than another?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, ethnicity, sex) • Other medications • Comorbidities (alcohol or other substance abuse, dementia or Alzheimer's disease, arthritis, cancer, diabetes, HIV/AIDS, multiple sclerosis, somatizing depression, stroke, vascular disease, chronic heart failure, CHD, post-MI, and vascular disease)
DYSPEPSIA	DERP did not report specifically on dyspepsia. We sued a DERP report on GERD instead.			
McDonagh et al., 2009[94]	KQ1: What is the comparative effectiveness of different PPIs in patients with symptoms of GERD?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: What is the comparative effectiveness of different proton pump inhibitors in treating peptic ulcer and NSAID-induced ulcer?	No	NA	NA
	KQ3: What is the comparative effectiveness of different proton pump inhibitors in preventing ulcer in patients taking an NSAID?	• Co-interventions	Not stated explicitly	• Patients taking an NSAID
	KQ4: What is the comparative effectiveness of different proton pump inhibitors in eradicating *Helicobacter pylori* infection?	No	NA	NA
	KQ5: Is there evidence that a particular treatment strategy is more effective and safer than another (for example, stepping down to a lower dose, treatment as needed compared with daily treatment, high dose compared with standard dose, or switching to an H2 antagonist) for treatment longer than 8 weeks in patients with gastroesophageal reflux disease or ulcer?	No	NA	NA
	KQ6: What are the comparative safety and adverse events of different PPIs in patients being treated for symptoms of gastroesophageal reflux disease, peptic ulcer, and NSAID-induced ulcer?	No	NA	NA
Evidence	KQ7: Are there subgroups of patients based on demographics, other medications, or comorbidities (including patients with nasogastric tubes, or who cannot swallow solid oral medications) for which a particular proton pump inhibitor or preparation is more effective or associated with fewer adverse effects?	• Demographics, unspecified • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, gender, racial groups) • Other medications • Comorbidities (e.g., nasogastric tubes, cannot swallow solid oral meds)
HYPERTENSION	NOTE: The DERP hypertension reports are two separate drug glass reports with different authors. On the DERP website, one report falls under “ACE Inhibitors” and the other under “Angiotensin II Receptor Antagonists” Both reports use the same key questions, and both had their key questions updated in 2009. Thus, we are using the most recent (and identical) questions.			

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Chou et al., 2005[88]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with ACE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ2: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy what are the inter-class differences in effectiveness and efficacy between DRI, ACE-I and AIIRA drugs? 2a. When used as monotherapy? 2b. When used in combination with one another?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ3: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in harms between DRI, ACE-I and AIIRA drugs?	No	NA	NA
	KQ4: Are there subgroups based on demographics (age, racial groups, gender), other medications, or co-morbidities for which there are inter-class differences between DRI, ACE-I and AIIRA drugs?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities
Furmaga et al., 2006[89]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with ACE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in effectiveness and efficacy between DRI, ACE-I and AIIRA drugs? 2a. When used as monotherapy? 2b. When used in combination with another one?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ3: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in harms between DRI, ACE-I and AIIRA drugs?	No	NA	NA
	KQ4: Are there subgroups based on demographics (age, racial groups, gender), other medications, or co-morbidities for which there are inter-class differences between DRI, ACE-I and AIIRA drugs?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities
Helfand et al., 2009[90]	KQ1: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine or bleeding esophageal varices, do beta blocker drugs differ in effectiveness/efficacy?	No	NA	NA
	KQ2: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine prophylaxis or bleeding esophageal varices, do beta blocker drugs differ in harms?	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, gender), other medications (drug-drug interactions), or co-morbidities (drug-disease interactions) for which one beta blocker is more effective or associated with fewer adverse effects?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications (drug-drug interactions) • Comorbidities (drug-disease interactions)

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
McDonagh et al., 2005[95]	KQ1: Do CCBs differ in effectiveness in the treatment of adult patients with essential hypertension (blood pressure ≥ 140/90 mm Hg), angina, supraventricular arrhythmias, or systolic dysfunction (LVEF <45%)?	No	NA	NA
	KQ2: Do CCBs differ in their safety or adverse effects in the treatment of adult patients with essential hypertension (blood pressure ≥ 140/90 mm Hg), angina, supraventricular arrythmias, or systolic dysfunction (LVEF <45%)?	No	NA	NA
	KQ3: Based on demographics (age, racial groups, gender), other medications, or co-morbidities, are there subgroups of patients for which one CCB is more effective or is associated with fewer adverse effects?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Co-morbidities
IRRITABLE BOWEL SYNDROME	DERP did not produce a report on irritable bowel syndrome. Thus, we used a DERP report on constipation			
Gartlehner et al., 2007[96]	KQ1: What is the general effectiveness of drugs used to treat chronic constipation associated with IBS? Given general effectiveness, what is the comparative effectiveness of drugs used to treat chronic constipation and chronic constipation associated with IBS?	No	NA	NA
	KQ2: Does treatment duration influence the effectiveness of drugs used to treat chronic constipation and chronic constipation associated with IBS? When should treatments be switched in patients not responding to a given drug?	No	NA	NA
	KQ3: What is the comparative tolerability and safety of drugs used to treat chronic constipation and chronic constipation associated with IBS?	No	NA	NA
	KQ4: Are there subgroups of patients based on demographics (age, racial or ethnic groups, and gender), other medications, or co-morbidities, including IBS, for which one symptomatic treatment is more effective or associated with fewer adverse events?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial or ethnic groups, and gender) • Other medications • Comorbidities (including IBS)

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
MYOCARDIAL INFARCTION				
Chou et al., 2005[88]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with CAE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ2: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in effectiveness and efficacy between DRI, ACE-I and AIIRA drugs? 2a. When used as monotherapy? 2b. When used in combination with another one?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ3: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what are the inter-class differences in harms between DRI, ACE-I and AIIRA drugs?	No	NA	NA
	KQ3: Are there subgroups based on demographics (age, racial groups, gender), other medications, or co-morbidities for which there are inter-class differences between DRI, ACE-I and AIIRA drugs?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Furmaga et al., 2006[89]	KQ1: For adults with diagnosed coronary heart disease, hypertension, left ventricular dysfunction, heart failure, nondiabetic chronic kidney disease, or diabetic nephropathy, what is the effectiveness and efficacy and what are the harms of aliskiren compared with placebo? 1a. When used as monotherapy? 1b. When used in combination with ACE-I and AIIRA drugs?	• Co-interventions	Not stated explicitly	• Age (adults18+ years)
	KQ2: For adult patients with essential hypertension, high cardiovascular risk factors, recent myocardial infarction, heart failure, diabetic or nondiabetic nephropathy, do angiotensin II receptor antagonists differ in safety or adverse events? The outcomes of interest with regard to safety include: a. Overall adverse effect reports b. Withdrawals due to adverse effects c. Serious adverse events reported (including mortality) d. Specific adverse effects or withdrawals due to specific adverse events (e.g., renal impairment, cough, and angioedema)	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, gender), other medications, or co-morbidities for which one angiotensin II receptor antagonist is more effective or associated with fewer adverse events (e.g., renal insufficiency)? Evidence unique to minority and ethnic groups are of particular interest.	• Age, race, ethnicity, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities (including diabetes)
Dailey et al., 2007[97]	KQ1: For adult patients with acute coronary syndromes or coronary revascularization via stenting or bypass grafting, prior ischemic stroke or transient ischemic attack, or symptomatic peripheral vascular disease, do antiplatelet drugs differ in effectiveness?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ2: For adults with acute coronary syndromes or coronary revascularization via stenting or bypass grafting, prior ischemic stroke or transient ischemic attack, or symptomatic peripheral vascular disease, do antiplatelet drugs differ in safety or adverse events?	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, gender), other medications (drug-drug interactions), comorbidities (drug-disease interactions), or pregnancy for which a particular antiplatelet drug is more effective or associated with fewer adverse events?	• Age, race, sex • Co-interventions • Comorbidities • Pregnancy	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities (diabetes, PAD, IHD, previous cardiac surgery, pre-existing atherosclerotic disease) • Pregnancy
Helfand et al., 2006[90]	KQ1: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine or bleeding esophageal varices, do beta blocker drugs differ in effectiveness/efficacy?	No	NA	NA
	KQ2: For adult patients with hypertension, angina, coronary artery bypass graft, recent myocardial infarction, heart failure, atrial arrhythmia, migraine prophylaxis or bleeding esophageal varices, do beta blocker drugs differ in harms?	No	NA	NA
	KQ3: Are there subgroups of patients based on demographics (age, racial groups, gender), other medications (drug-drug interactions), or co-morbidities (drug-disease interactions) for which one beta blocker is more effective or associated with fewer adverse effects?	• Age, race, sex • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications • Comorbidities

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Helfand et al., 2006[98]	KQ1: How do statins and fixed-dose combination products containing a statin and another lipid lowering drug compare in their ability to reduce LDL-c? Are there doses for each statin or fixed-dose combination product containing a statin and another lipid lowering drug that produce similar percent reduction in LDL-c between statins? Is there a difference in the ability of a statin or fixed-dose combination product containing a statin and another lipid lowering drug to achieve National Cholesterol Education Panel goals?	No	NA	• Age (>18 years)
	KQ2: How do statins and fixed-dose combination products containing a statin and another lipid lowering drug compare in their ability to raise HDL-c? Are there doses for each statin or fixed-dose combination product containing a statin and another lipid lowering drug that produce similar percent increase in HDL-c between statins? Is there a difference in the ability of a statin or fixed-dose combination product containing a statin and another lipid lowering drug to achieve National Cholesterol Education Panel goals?	No	NA	• Age (>18 years)
	KQ3: How do statins and fixed-dose combination products containing a statin and another lipid lowering drug compare in their ability to reduce the risk of nonfatal myocardial infarction, coronary heart disease (angina), coronary heart disease mortality, all-cause mortality, stroke, hospitalization for unstable angina, or need for revascularization (coronary artery bypass graft, angioplasty, or stenting)?	No	NA	• Age (>18 years)
	KQ4: Are there differences in the effectiveness of statins and fixed-dose combination products containing a statin and another lipid lowering drug in different demographic groups or in patients with comorbid conditions (e.g., diabetes, obesity)?	• Demographics, unspecified • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Comorbid conditions (e.g., diabetes, obesity)

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ5: Are there differences in the harms of statins or fixed-dose combination products containing a statin and another lipid lowering drug when used in the general population of children or adults?	Age	Not stated explicitly	• Age (children or adults)
	KQ6: Are there differences in the harms of statins or fixed-dose combination products containing a statin and another lipid lowering drug when used in special populations or with other medications (drug-drug interactions)? In addressing this question, we will focus on the following populations: • Patients with HIV • Organ transplant recipients • Patients at high risk for myotoxicity (e.g., patients with a history of statin-associated muscle-related harms due to drug-drug/drug-food interactions, patients co-administered fibrates, patients taking potent 3A4 inhibitors, elderly patients, especially elderly females) • Patients at high risk for hepatotoxicity • Patients using fibrates (gemfibrozil, fenofibrate, fenofibric acid) or niacin • Children with nephritic syndrome	• Risk factors • Age, sex • Co-interventions • Comorbidities	Not stated explicitly	• Patients with HIV • Organ transplant recipients • Patients at high risk for myotoxicity • Patients at high risk for hepatotoxicitly • Patients using fibrates • Children with nephritic syndrome
OSTEOARTHRITIS				
Chou et al., 2006[99]	KQ1: Are there differences in effectiveness between coxibs and other NSAIDs?	No	NA	NA
	KQ2: Are there clinically important differences in short-term safety or adverse effects between coxibs, other NSAIDs, and the combination of an NSAID plus antiulcer medication when used for musculoskeletal pain?	No	NA	NA
	KQ3: Are there clinically important differences in long-term safety or adverse effects between coxibs, other NSAIDs, and the combination of an NSAID plus antiulcer medication when used chronically?	No	NA	NA

Table C5. Drug Effectiveness Review Program (DERP) reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ4: Are there subgroups of patients based on demographics, other medications (e.g., aspirin), or co-morbidities for which one medication is more effective or associated with fewer adverse effects?	• Demographics, unspecified • Co-interventions • Comorbidities	Not stated explicitly	• Demographics (age, racial groups, gender) • Other medications (concomitant aspirin or anticoagulant use) • Comorbidities

Table C6. Health Technology Assessment (HTA) final reports

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
CANCER, BREAST				
Adelaide Health Technology Assessment, 2008[100]	Objective: To critically appraise and synthesis evidence about the effectiveness of breast cancer screening overall and among women younger than 50 years. Research questions: • What are the safety, effectiveness and cost-effectiveness of DM, compared with FM, for screening for breast cancer in asymptomatic women aged 40 years and over attending a breast cancer screening program? • What are the safety, effectiveness and cost-effectivness of DM + breast ultrasound scan + MRI, compared with FM + breast ultrasound scan + MRI, for surveillance of breast cancer in women who are at potentially high risk of breast cancer? • What are the safety, effectiveness and cost-effectiveness of DM + breast ultrasound scan + MRI, compared with FM + breast ultrasound scan + MRI, for diagnosing breast cancer in women presenting with signs or symptoms of malignancy?	• Age	Not stated explicitly	• Age • Density of breast tissue Inclusion Criteria: • Q1. women aged 40+ years • Q2. Women at potentially high risk of breast cancer • Q3. Women presenting with signs or symptoms of malignancy

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Korencan et al., 2007[101]	Policy question: Are resources allocated to treat HER2 patients being used in the most efficient way? Research question: What is the gold standard for diagnosing HER2 positive tumors? Which is the most accurate and reproducible method to identify candidates for potential therapy with monoclonal antibodies, and are the applied tests reliable for selecting HER2 positive patients? Is it necessary to look closer at specific areas of uncertainty—if so, which areas?	No	NA	Included studies examines some/all of the following: • Age • Tumor grade • Nuclear score • Tubule score • Mitotic score • Ki-67 % • ER % • PR % Exclusion* criteria: • HER2 expression in tumours other than mamma • Male mamma carcinoma • Radiological diagnostics of carcinoma: HER2 vs. non-HER2 • Monitoring response to chemotherapeutic schemes • Correlation between HER2 status and worse outcome • Special subpopulations (pregnant) *Due to wording of inclusion vs. exclusion criteria, including exclusion criteria made more sense

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Dunfield and Severn, 2007[102]	Research Questions: 1. What is the clinical effectiveness of MRI screening compared to film mammography in women with a high risk of breast cancer? 2. What is the cost-effectiveness of MRI screening compared to film mammography in women with a high risk of breast cancer? 3. What is the strength of evidence used to support the American Cancer Society's guidelines regarding MRI screening for women at high risk of breast cancer?	• Risk factors	Based on National Cancer Institute Gail and BRCAPRO models	• Age • Race • Incidence of breast cancer in first-degree relatives • Number of times that a woman has given birth (parity) Age at menarche • Number of breast biopsies • Atypical hyperplasia Inclusion criteria: • Women at high risk of breast cancer.
CANCER, LUNG				
Adelaide Health Technology Assessment, 2007[103]	Objective: Examine recent evidence for the use of CT utilized for lung cancer screening, and its applicability to the Australian setting.	No	NA	• Non-small cell lung cancer (including squamous carcinoma, adenocarcinoma, and large cell carcinoma) vs. small cell lung cancer • Current vs. former smokers
CANCER, PROSTATE				
Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen, 2007[104]	Research question: Evaluate the benefits and harms of low-dose-rate permanent interstitial brachytherapy in localised prostate cancer compared with standard surgical procedures, percutaneous radiotherapy, and watchful waiting. The focus of the evaluation was on patient-relevant therapy goals. Moreover, substantially different types of brachytherapy were to be compared with each other.	• Severity, stage, or site	Not stated explicitly	Inclusion criteria: • Men with localized prostate cancer • Age • Baseline PSA values • Gleason Score • Clinical tumour stage

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Pearson et al., 2007[105]	KQ1: What are the effects of IMRT vs. 3D-CRT on survival, disease-free survival, incidence of adverse side effects, quality of life, and health care utilization and costs? KQ2: Based on these findings, what is the estimated cost per serious adverse event prevented and the cost per quality adjusted life-year gained for IMRT vs. 3D-CRT? KQ3: What are the key patient clinical characteristics that may influence the clinical and cost-effectiveness of IMRT compared to 3D-CRT?	• Unspecified subgroups	Not stated explicitly	Inclusion criteria: • Men with localized prostate cancer. This was largely a cost-effectiveness study. The only clinical characteristic mentioned was "clinicians can estimate higher risks for patients with certain underlying characteristics, such as history of inflammatory bowel disease or anticoagulation use," although no estimates were provided.
DEPRESSION				
Swedish Council on Technology Assessment in Health Care, 2007[106]	KQ1: Is light therapy more effective than placebo in treating SAD?	No	NA	NA
	KQ2: Are there other effective (drug) therapies for SAD?	No	NA	NA
	KQ3: Does the effect of antidepressants on non-seasonal depression set in more rapidly when light therapy is used as an adjunct treatment?	No	NA	NA
	KQ4: Is light therapy a more effective monotherapy than placebo in treating non-seasonal depression?	No	NA	NA
Swedish Council on Technology Assessment in Health Care, 2007[107]	KQ: What effects and costs are associated with computer-based CBT in treating adult patients with anxiety disorders or depression?	No	NA	NA
McLoughlin et al., 2007[108]	Objective: To investigate if rTMS was as effective as ECT in treating major depressive episodes and to perform a cost-effectiveness analysis	No	NA	Inclusion criteria: • Right-handed patients (age 18+ years) with major depressive episodes referred for ECT.

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Screening for postnatal depression within the Well Child Tamariki Ora Framework, 2008[109]	Objective: To evaluate the potential value for money of implementing a screening programme for postnatal depression within the Well Child Tamariki Ora Framework.	No	NA	Inclusion criteria: • New mothers in New Zealand who have given birth in any 12 month period.
DYSPEPSIA				
Swedish Council on Technology Assessment in Health Care, 2007[110]	KQ1: Do acid inhibitors relieve symptoms more effectively than placebo? KQ2: Does drug therapy relieve symptoms more effectively than placebo when there is coexisting H. pylori infection? KQ3: Are PPIs more cost-effective than placebo?	No	NA	NA
HYPERTENSION				
Hermanowski, et al., 2007[111]	Objective: To quantify the impact on the public payer's budget of reimbursement of the following medications: bosentan (Tracleer), epoprostenol (Flolan), iloprost (Ventavis), sildenafil (Revatio) and treprostinil (Remodulin) used in treatment of NYHA III/IV PAH	No	NA	NA
Hermanowski et al., 2007[112]	Objective: To evaluate prevalence and current practice of treatment of patients with NYHA III/IV pulmonary arterial hypertension in Poland as well as expected changes due to introduction of novel drugs: bosentan (Tracleer), epoprostneol (Flolan), iloprost (Ventavis), sildenafil (Revatio) and treprostinil (Remodulin)	No	NA	NA
Tran et al., 2007[113]	Objective: To evaluate the evidence for the clinical effects and the economic implications of thiazide diuretics when used as a first-line treatment for hypertension.	No	NA	• Age • Gender • Race • Co-morbid conditions • Baseline severity of hypertension

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
MYOCARDIAL INFARCTION				
Agence d'Evaluation des Technologies et des modes d'Intervention en Sante, 2008[114]	Primary Objective: To compare the efficacy (using meta-analyses of RCTs) and effectiveness (using meta-analysis of observational studies) of PPCI and FL, in terms of mortality reduction. Secondary Objectives • Efficacy • To compare the incidences of reinfarction associated with PPCI and FL • Safety • To compare the incidences of stroke associated with PPCI and FL • To compare the incidences of major bleeds associated with PPCI and FL	No	NA	Authors noted that a lack of individual patient data prevented subgroup comparison (e.g., elderly, high-risk STEMI) and was a limitation to the study.
National Coordinating Center for Health Technology Assessment, 2007[115]	KQ: The study aimed to evaluate the relative effectiveness and cost-effectiveness of a home-based programme of cardiac-rehabilitation using the Heart Manual, with centre-based programmes in patients who have experienced a MI or coronary revascularization within the previous 12 weeks. In addition, it sought to explore the reasons for non-adherence to cardiac rehabilitation programmes.	No	NA	• Age • Sex • Ethnicity • Recruitment diagnosis • Previous CR experience Inclusion criteria: • Adult patients who have experienced a myocardial infarction or coronary revascularization within the previous 12 weeks.
National Institute for Health and Clinical Excellence, 2007[116]	*Duplicated under NICE (and randomly selected twice)*			
OSTEOARTHRITIS				
Canadian Agency for Drugs and Technologies in Health, 2007[117]	Objective: To examine the evidence regarding the use of hyaluronic acid or hyaluronan for hip OA and complements an earlier bulletin on HA for knee OA.	No	NA	• NA

Evidence Table C6. Health Technology Assessment (HTA) final reports (continued)

Reference	Key Question(s)	CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	CH or subgroups addressed in analysis? If yes, how has CH been addressed?
Samson et al., 2007[118]	KQ1: What are the clinical effectiveness and harms of each intervention in patients with primary OA of the knee?	• Risk factors • Severity, stage, or site	Not stated explicitly	• Patients with primary OA of the knee
	KQ2: What are the clinical effectiveness and harms of each intervention in patients with secondary OA of the knee?	• Risk factors • Severity, stage, or site	Not stated explicitly	• Patients with secondary OA of the knee
	KQ3: How do the short-term and long-term outcomes of each intervention differ by the following subpopulations: age, race/ethnicity, gender, primary or secondary OA, disease severity and duration (body mass index), and prior treatments?	• Risk factors • Age, race, ethnicity, sex • Severity, stage, or site	Not stated explicitly	• Age • Race/ethnicity • Gender • Primary or secondary OA • Disease severity and duration • Body weight and/or BMI • Prior treatments
	KQ4: How do the short-term and long-term outcomes of each intervention compare for the treatment of primary OA of the knee; and secondary OA of the knee?	• Risk factors • Severity, stage, or site	Not stated explicitly	• Primary vs. secondary OA of the knee

Table C7. National Institute for Health and Clinical Excellence (NICE)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
DEPRESSION				
National Collaborating Centre for Mental Health[119]			Clinical questions were used to guide the identification and interrogation of the evidence base. The questions were developed using a modified nominal group technique. The process began by asking each member of the GDG to submit as many questions as possible. The questions were then collated and refined by the review team. At a subsequent meeting, the guideline chair facilitated a discussion to further refine the questions. At this point, the GDG members were asked to rate each question for importance. The results of this process were then discussed and consensus reached about which questions would be of primary importance and which would be secondary. The GDG aimed to address all primary questions, while secondary questions would only be covered time permitting.	

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
A. SERVICE TOPIC GROUP				
	KQ1. Does screening for depression by GPs improve outcomes?	No		No; Not in E.S., however, recommendation 5.2.4.1 and clinical summary (p77) denote high risk groups for depression – those with a past history; significant physical illnesses causing disability or other mental health problems such as dementia; women around the time of child birth; those with chronic drug or alcohol abuse
	KQ2. In depression, does guided self-help improve outcomes compared to other interventions?	No		Clinical summary 5.3.5: patients with mild to moderate depression
	KQ3. Does computerized CBT improve patient outcomes compared to other treatments?	No		Yes; clinical summary 5.4.5: mild and moderate depression
	KQ4. Does exercise improve patient outcomes compared to other treatments or TAU?	No		Yes; ES 5.5.4.5: community-dwelling depressed older individuals Clinical summary 5.5.5: patients with mild/moderate depressive disorder, individuals with low mood
	KQ5. In depression, which model of care produces the best outcomes?	No		No
	KQ6. Do non-statutory support groups improve outcomes?	No		Yes; clinical summary 5.7.5: women with chronic depression
	KQ7. Do crisis resolution and home treatment teams improve patient outcomes compared to other treatments?	No		Yes; clinical summary 5.8.5: pts with depression that require a higher level of care than can be provided by standard community services

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ8. Do day hospitals improve patient outcomes compared to other treatments?	No		Yes, ES 5.9.1.4: people with a diagnosis of mood or anxiety disorder P109: Studies predominantly examined patients 18-65 years old. studies excluded pts that were predominantly either over 65 years or under 18 years of age
	KQ9. Does electroconvulsive therapy improve patient outcomes compared to other treatments?	No		No
B. PSYCHOLOGY TOPIC GROUP				
	KQ10. Are psychological interventions effective compared to: • treatment as usual • other psychological interventions • medication	No		Yes, 6.1.5 (p119): older adults with depression
	KQ11. Is there a benefit in combining psychological interventions with medication?	No		Yes, ES 6.9.3.2 older pts

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
C. PHARMACOLOGY TOPIC GROUP				
National Collaborating Centre for Mental Health[119]	KQ12. Is any single (or class of) antidepressant better in the treatment of depression?	No		Yes; ES 7.7.3: moderate/severe/very severe depression Clinical summary 8.3.1.6: women, those with suicidal intent Clinical summary 8.1.4.3.4: pts with MDD Clinical Summary 8.1.6.5: moderate and severe depression
	KQ13. Does the choice of antidepressant depend on: • Severity of depression (including threshold) • Severity of depression (including threshold) • Depression sub-type (psychotic depression or depression with atypical features) • Side effects • Discontinuation symptoms • Setting • Gender • Age • Setting	• Risk factors • Age, sex • Severity, stage, or site • Comorbidities		Yes; ES 6.2.4.10: people with up to two episodes of depression; people who have had more than two episodes of depression ES 8.2.2.2.2: older adults ES 8.2.3.4: gender ES 8.2.4.3: pts with psychotic depression ES 8.2.5.3: pts with atypical depression ES 8.2.6.3: pts with multiple relapses of depressive episodes
	KQ14. What pharmacological strategies are effective in refractory depression?	• Severity, stage, or site		Yes; p240: people with treatment-resistant depression (those whose depression symptoms failed to respond to two or more antidepressants at an adequate dose for an adequate duration given sequentially); acute phase non-responder: participants who failed only one course of antidepressants

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ15. Is St John's wort effective in depression? • By severity of depression • Compared to antidepressants	• Severity, stage, or site		Yes; ES 8.1.6.3: moderate, severe depression
	KQ16. Which switching strategies are effective?	No		No
	KQ17. What are the best pharmacological management strategies to prevent relapse?	No		Yes; ES 8.2.6.3: pts that achieved remission while taking an antidepressant plus lithium
DYSPEPSIA				
North of England Dyspepsia Guideline Development Group, 2004[120]	KQ1. How is dyspepsia defined; what is and what isn't dyspepsia?	No		
	KQ2. What is the appropriate role of the pharmacist in managing dyspepsia?	No		
	KQ3. How should dyspepsia be diagnosed in primary care?	No		
	KQ4. How can dyspepsia in primary care be characterized in terms of its presentation, psychological influence, and impact upon patient quality-of-life?	No		
	KQ5. What factors prompt patients to consult for dyspepsia?	Yes; unspecified		
	KQ6. How should symptoms be assessed and interpreted?	No		
	KQ7. How should diagnosis be organized?	No		
	KQ8. How should dyspepsia be managed in primary care?	No		
	KQ9. How can communication be promoted, embracing patient expectation and promoting understanding?	No		

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ10. Do lifestyle interventions work?	No		
	KQ11. Which acid suppressing therapy should be used and for how long?	No		
	KQ12. Who should get H. Pylori eradication therapy and with which regimen?	Yes; unspecified		
	KQ13. What is the relationship between NSAID therapy and dyspepsia?	Yes; unspecified		
	KQ14. How should long term care be organized in its frequency and content and with regard to patient safety?	No		
	KQ15. What are appropriate grounds for referral?	Yes; unspecified		
	KQ16. What are the risks of serious underlying pathology?	• Risk factors • Comorbidities		
	KQ17. How should these risks be conceptualized and discussed by clinicians and patients?	• Risk factors • Comorbidities		
	KQ18. What are alarm signals and what should be done when they occur?	• Risk factors • Comorbidities		
MYOCARDIAL INFARCTION				
Cooper et al., 2007[121]			**The methodology team and the GDG agreed that a full literature search should not be undertaken for all 55 MI KCQs. The methodology team and GDG identified the KCQs requiring a full literature search and critical appraisal. Lit searches were not undertaken where there was already national guidance on the topic to which the guideline could cross reference.	

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ1. What is the effectiveness of changing dietary regime from the pre-infarct diet?	No–the guidelines are geared towards adults (≥18 years) who have had an MI and denotes two clinical pathways: 1) patients following the early acute phase (48 hours after admission provided the patient is stable) and 2) patients identified as having had a proven MI at some point in the past.		Yes; Evidence Statements (E.S.) 4.2.1.7 (pts with hypercholesterolemia after an MI)
	KQ2. What education and/or information best aids patients after MI to (i) reduce their risk of subsequent cardiac problems (ii) return to a full and normal life (daily activities, driving, exercise, employment, leisure activities, sexual activities)	No		No
	KQ3. What psychological and social (careers) support best aids people after MI to reduce their risk of subsequent cardiac problems and to promote their return to a full and normal life?	No		No
	KQ4. What is the incidence of sexual dysfunction in patients after MI and how can patients be identified who would require referral to a specialist unit?	No		Yes; E.S. 5.6.2.1: references male pts after MI with ED

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ5. What is the effectiveness of adding ACEI versus placebo to improve outcome in i) unselected patients after MI? ii) patients after MI with LV dysfunction?	• Comorbidities		Yes; E.S. 6.2.1.5: pts after MI with heart failure or LVSD
	KQ6. What is the effectiveness of adding ARBs versus placebo to improve outcome in (i) patients after MI without LV dysfunction? (ii) patients after MI with LV dysfunction?	• Comorbidities		Yes; E.S.6.2.1.7: pts after acute MI
	KQ7. What is the effectiveness of adding ACEI versus ARBs to improve outcome in (i) unselected patients after MI? (ii) patients after MI with LV dysfunction?	• Comorbidities		Yes; E.S. 6.2.1.10: pts early after MI without HF or LVSD. E.S.6.2.1.11: pts with HF and/or LVSD treated within 10 days of acute MI
	KQ8. What is the effectiveness of adding ACEI plus ARBs versus ACEI to improve outcome in patients after MI with LV dysfunction?	• Comorbidities		Yes; E.S. 6.2.1.13: pts with HF and/or LVSD treated within 10 days of acute MI
	KQ9. How frequently should renal function tests, including serum potassium, be monitored in patients treated with ACEI and/or ARBs after MI?	• Co-interventions		Yes
	KQ10. What is the effectiveness of adding aspirin versus placebo to improve outcome in patients after MI?	No		No
	KQ11. What is the effectiveness of adding aspirin versus clopidogrel to improve outcome in patients after MI?	No		Yes; E.S. 6.3.1.2: pts with recent MI
	KQ12. What is the most effective method of delivering dietary advice?	No		No
	KQ13. What is the effectiveness of adding aspirin versus aspirin and clopidogrel to improve outcome in…	• Severity, stage, or site		Yes; E.S. 6.3.1.5: Pts with an ST segment elevation acute coronary syndrome

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ14. What is the effectiveness of adding a beta blocker versus placebo to improve outcome in...	• Comorbidities		Yes; E.S. 6.4.1.1: pts after acute MI E.S. 6.4.1.3: pts with LV dysfunction
	KQ15. What is the effectiveness of adding vitamin K antagonist (warfarin) versus placebo to improve outcome in patients after an MI?	No		Yes; E.S. 6.5.1.1: pts after acute MI
	KQ16. What is the effectiveness of adding vitamin K antagonist (warfarin) versus aspirin to improve outcome in patients after an MI?	No		Yes; E.S. 6.5.1.4: Pts with CAD
	KQ17. What is the effectiveness of adding vitamin K antagonist (warfarin) plus aspirin versus aspirin to improve outcome in patients after MI?	No		Yes; E.S. 6.5.1.5: pts after acute MI E.S. 6.5.1.6: pts after acute MI E.S. 6.5.1.7: pts after acute MI
	KQ18. What is the effectiveness of adding vitamin K antagonist (warfarin) plus aspirin versus warfarin to improve outcome in patients after MI?	No		Yes; E.S. 6.5.1.8: pts after acute MI
	KQ19. What is the effectiveness of adding calcium channel blocker versus placebo to improve outcome in…	• Comorbidities		No
	KQ20. What is the effectiveness of adding potassium channel activators versus placebo to improve outcome in patients after MI?	No		No
	KQ21. How frequently should renal function, including serum potassium, be monitored in patients post MI treated with eplerenone?	No		No
	KQ22. What is the effectiveness of adding eplerenone versus placebo to improve outcome in patients after MI ?	No		No

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ23. What is the effectiveness of adding Omega-3-acid ethyl esters versus placebo to improve outcome in patients after MI?	No		No
KQ 24 and 26 are the same	KQ24. What is the effectiveness of low/ moderate alcohol consumption versus high alcohol consumption to improve outcome in patients after MI?	No		Yes; E.S. 4.4.1.1: consumption by men after an MI E.S. 4.4.1.2: consumption by women after an MI
	KQ25. What is the effectiveness of no/ low/moderate alcohol consumption versus high alcohol consumption to improve outcome in patients after MI?	No		No
	KQ26. What is the effectiveness of low/ moderate alcohol consumption versus high alcohol consumption to improve outcome in patients after MI?	No		Yes; E.S. 4.4.1.1: consumption by men after an MI E.S. 4.4.1.2: consumption by women after an MI
	KQ27. What is the effectiveness of adding statins versus placebo to improve outcome in patients after MI?	No		No
	KQ28. What is the effectiveness of adding high dose statin (more potent cholesterol lowering) versus low dose statin (less potent cholesterol lowering) to improve outcome in patients after MI?	No		No
	KQ29. What is the effectiveness of adding early statin therapy versus delayed statin therapy to improve outcome in patients after MI?	No		No
	KQ30. What is the effectiveness of adding fibrates or niacin or ezetimibe versus placebo to improve outcome in patients after MI?	No		No

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ31. Are there stable patients who don't benefit prognostically from revascularization	• Severity, stage, or site		No
	KQ32. Are there stable patients after MI who a) benefit prognostically from revascularisation b) those who don't benefit prognostically	• Severity, stage, or site		No
	KQ33. What is the optimal target blood pressure for patients after MI with hypertension? Assuming a patient is treated with ACEI and or ARB and a beta blocker already (and in LV dysfunction and HF eplerenone)	• Co-interventions		No
	KQ34. Does determining LV function versus standard care improve (that is, affect) outcome of patients MI (summarising LV dysfunction effect on drugs/ ICD /rehab)?	No		No
	KQ35. Is there any benefit in giving ACEI at a later stage of treatment in patients with previous MI (later than one year)	• Severity, stage, or site		No
KQ 36 and KQ 46 are the same	KQ36. Does a history of proven MI in the past (> 1 year) versus recent MI (< 1 year) change treatment / management / outcome?	• Severity, stage, or site		Yes; pts with past MI vs recent MI
KQ 37 and KQ 45 are the same	KQ37. What is the effectiveness of regular physical activity versus a sedentary lifestyle to improve outcome in patients after MI?	No		No
	KQ38. What is the level of physical activity which increases physical work capacity versus physical activity which does not increase physical work capacity	No		No
	KQ39. What is the effectiveness of comprehensive cardiac rehabilitation versus standard care with no cardiac rehabilitation to improve outcome in patients after MI ?	No		No

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ40. What is the effectiveness of exercise only cardiac rehabilitation versus standard care with no cardiac rehabilitation to improve outcome in patients after MI?	No		Yes; E.S. 5.2.1.6: stable patients with LV dysfunction E.S. 5.2.1.7: older people
	KQ41. What is the effectiveness of comprehensive cardiac rehabilitation versus exercise only cardiac rehabilitation to improve outcome in patients after MI?	No		No
	KQ42. What is the effectiveness of an individualised cardiac rehabilitation programme versus a non-individualised cardiac programme to improve outcome in patients after MI?	No		No
	KQ43. Are there any patients after MI in whom the exercise component of cardiac rehabilitation is not safe?	No		No
	KQ44. What approach to patient engagement best aids access to cardiac rehabilitation, particularly in reference to em, op, seg, women, those from rural communities, and those with mental and physical health co-morbidities?	• Risk factors • Age, ethnicity • Comorbidities		E.S. 5.3.1.6 cited that there was no evidence of interventions to improve uptake or adherence to cardiac rehab in em, op, seg, women, those from rural communities. The authors attempted to address CH in their analyses, but were unable to do so due to lack of studies meeting the inclusion/exclusion criteria for the review.
	KQ45. What is the effectiveness of regular physical activity versus a sedentary lifestyle to improve outcome in patients after MI?	No		No
	KQ46. Does a history of proven MI in the past (> 1 year) versus recent MI (< 1 year) change treatment / management / outcome?	• Severity, stage, or site		Yes; pts with past MI vs recent MI

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ47. What is the effectiveness of adding fibrates versus placebo to improve outcome in patients with CHD	No		No
	KQ48. What is the effectiveness of adding ezetimibe versus placebo to improve outcome in patients with CHD	No		No
	KQ49. Is there an optimum time for ACEI to be administered in the nonacute phase?	No		No
	KQ50. Is there and optimum time for beta-blockers to be initiated in unselected patients after MI?	No		Yes; E.S. 6.4.1.1: unselected pts after acute MI
	KQ51. What is the potential harm of adding the following. calcium channel blocker or thiazide diuretic or alpha blocker versus placebo in…	• Comorbidities		No
	KQ52. What is the incidence of anxiety and depression in patients after MI and how can patients be identified? (can be cross-referenced to the Anxiety & Depression guidelines)	No		No
	KQ53. What are the information and support needs for patients at different points in the care pathway?	No		No
	KQ54. At what level of renal function do the risks of therapy with ACEIs outweigh the benefits in patients after MI with poor renal function	• Severity, stage, or site • Comorbidities		The authors attempted to address CH in their analyses, but were unable to do so due to lack of studies meeting the inclusion/exclusion criteria for the review.
	KQ55. Is there any benefit in initiating beta blockers at a later stage of treatment	No		No

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
OSTEOARTHRITIS				
National Collaborating Centre for Chronic Conditions[122]			"The technical team drafted a series of clinical questions that covered the guideline scope. The GDG and Project Executive refined and approved these questions."	
	KQ1. In adults with osteoarthritis, what are the benefits and harms of paracetamol compared with oral and NSAIDs or selective COX-2 inhibitors with respect to pain reduction?	No		Yes; Table 7.1: pain outcomes
	KQ2. In adults with osteoarthritis, what are the benefits and harms of paracetamol alone compared with and opioids alone or ii) paracetamol-opioid compounds with respect to pain reduction?	No		Yes • Table 7.9: pain outcomes • Table 7.10: function outcomes • Table 7.11: stiffness outcomes
	KQ3. In adults with osteoarthritis, what are the benefits, and harms of paracetamol-opioid compounds and compared with NSAIDs with respect to pain reduction?	No		No
	KQ4. In adults with osteoarthritis, what are the benefits, and harms of low dose opioids with or without and paracetamol versus higher strength opioids with respect to pain reduction?	No		Yes • Table 7.15: OA site (knee and/or hip); pain outcomes • Table 7.16: stiffness outcomes • Table 7.17: function outcomes
	KQ5. In adults with osteoarthritis, what are the benefits, and harms of paracetamol compared with placebo and with respect to pain reduction?	No		Yes • Table 7.22: site of disease (knee and/or hip); pain outcomes • Table 7.23: stiffness outcomes

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ6. In adults with osteoarthritis, what are the benefits, and harms of tricyclics/SSRI/SNRI drugs versus and placebo with respect to symptoms, function and quality of life?	No		Yes: ES 7.1.11: function outcomes
	KQ7. In adults with osteoarthritis, what are the benefits, and harms of COX-2 inhibitors compared to and nonselective NSAIDs or ii) placebo with respect to symptoms, function and quality of life?	No		Yes; E.S. 7.3.3: OA site (knee, hip, hand, foot, knee and hip, mixed sites); pain outcomes;
	KQ8. In adults with osteoarthritis, what are the relative, benefits and harms of i) selective COX-2 inhibitors and versus nonselective NSAIDs plus GI protective agents and ii) selective COX-2 inhibitors plus GI protective agents versus nonselective NSAIDs plus GI protective agents?	No		No
	KQ9. In adults with osteoarthritis taking aspirin what are, the relative benefits and harms of selective COX-2 RCT and inhibitors versus nonselective NSAIDs versus each of these combined with GI protective agents?	No		No
	KQ10. In adults with osteoarthritis, what are the benefits and harms of topical agents (NSAIDs/capsaicin/ rubefacients) compared with oral NSAIDs or placebo with respect to symptoms, function and quality of life?	No		Yes • Table 7.30, Table 7.38, Table 7.46: OA site (knee and/or hand, and/or hip); pain outcomes • Table 7.39, Table 7.47: stiffness outcomes • Table 7.40, Table 7.48: function outcomes

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ11. In adults with osteoarthritis, what are the relative, benefits and harms of arthroscopic lavage (with or without debridement) versus i) tidal irrigation ii) sham procedure (placebo) with respect to symptoms, function and quality of life?	No		Yes • Table 6.104: OA site (knee); pain outcomes • Table 6.105: stiffness outcomes • Table 6.106: function outcomes
	KQ12. In adults with osteoarthritis, what are the relative, benefits and harms of intra-articular injection of and corticosteroid versus placebo with respect to symptoms, function, and quality of life?	No		Yes • Table 7.65: OA site (knee); pain outcomes • Table 7.67: function outcomes
	KQ13. In adults with osteoarthritis, what are the relative, benefits and harms of intra-articular injection of and hyaluronic acid/ hyaluronans versus placebo or steroid injection with respect to symptoms, function, and quality of life?	No		Yes; Table 7.72: OA site (knee, hip, hand); pain outcomes
	KQ14. In adults with osteoarthritis, what are the relative, benefits and harms of electrotherapy (ultrasound, and laser, transcutaneous electrical nerve stimulation [TENS, TNS, AL-TENS], pulsed shortwave diathermy, interferential therapy) versus no AMED 1985–2007 treatment, placebo or other interventions with respect to symptoms, function, and quality of life?	No		Yes • Table 6.36, Table 6.39: OA site (knee, thumb, hand); pain outcomes • Table 6.37, Table 6.47: function outcomes • Table 6.46: stiffness outcomes
	KQ15. In adults with osteoarthritis, what are the relative, benefits and harms of acupuncture versus sham and treatment (placebo) and other interventions with respect to symptoms, function, and quality of life?	No		Yes • Table 6.57: OA site (knee); pain outcomes • Table 6.59: function outcome

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ16. In adults with osteoarthritis, what are the relative, benefits and harms of glucosamine and chondroitin alone or in compound form versus placebo with respect to symptoms, function, and quality of life and ability to beneficiy modify structural changes AMED 1985–2007 of osteoarthritis?	No		Yes • Table 6.87: OA site (knee, hip, mixed); pain outcomes • Table 6.88: function outcomes
	KQ17. In adults with osteoarthritis, what are the relative, benefits and harms of local thermo-therapy (ice, and cold, warmth, hot packs, wax baths, contrast baths) versus no treatment or other interventions with respect to symptoms, function, and quality of life?	No		Yes • Table 5.10: OA site (knee); pain outcomes • Table 5.11: function outcomes
	KQ18. In adults with osteoarthritis, what are the relative benefits and harms of various manual therapies (massage, trigger point massage, mobilisation, manipulation) versus no treatment or other interventions with respect to symptoms, function, and quality of life?	No		Yes • Table 6.1: OA site (knee, hip); pain outcomes • Table 6.2, Table 6.8: function outcomes
	KQ19. In adults with osteoarthritis, what are the relative benefits and harms of rest and relaxation/application of pacing techniques versus no treatment or other interventions with respect to symptoms, function, and quality of life?	No		Yes; ES 5.3.3: OA site (hip, knee); pain outcomes
	KQ20. In adults with osteoarthritis, which devices (joint, brace, taping, strapping, splinting, footwear, insoles, and walking aids (cane, crutch, walker, walking stick, frame)) are the most effective when compared with one another or with no intervention/usual care with respect to symptoms, function, and quality of life?	No		Yes • Table 6.65: OA site (knee, thumb); pain outcomes • Table 6.67: function outcomes

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ21. In adults with osteoarthritis, are assistive devices (such as tap turners) more effective than no such devices in improving function and quality of life?	No		Yes • Table 6.74: OA site • Table 6.75: function outcomes
	KQ22. In adults with osteoarthritis, what are the indications for referring for consideration for total/partial joint replacement therapy?	No		Yes; E.S. 8.1.4: • OA site (knee, hip) • pt willingness to undergo surgery • structural features (destruction of joint space) • OA grade (severity) • postoperative care and physician advice
	KQ23. In adults with osteoarthritis, are there patient- centred factors that predict increased benefits or harms from osteoarthritis related surgery?	Yes; unspecified		Yes; E.S. 8.1.4: • OA site (knee, hip) • age • gender • weight/BMI • comorbidities • smoking/drugs/alcohol use • pain • pt willingness to undergo surgery • usage of assistive devices • pt psychological factors including expectations
	KQ24. In adults with osteoarthritis, what are the relative, benefits and harms of weight loss versus no weight and loss with respect to symptoms, function and quality of life?	No		Yes; Table 6.32: OA site (knee); pain outcomes

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ25. In adults with osteoarthritis, is exercise therapy more effective than i) placebo or no treatment or ii) other treatments (e.g., dietary, weight loss, education)?	No		Yes • Table 6.6: pain outcomes • Table 6.9: examination findings (e.g., mean peak torque values; walking time strength and range of knee flexion measures)
	KQ26. In adults with osteoarthritis, which type of exercise therapy is the most effective for reducing pain and disability?	No		Yes; Table 6.6: pain outcomes
	KQ27. In adults with osteoarthritis, what are the relative benefits of different patient information provision and/or education methods i) in relation to each other or ii) versus no specific information provision/ education, with respect to symptoms, function and quality of life?	No		Yes • Table 5.1: OA site (knee and/or hip); pain outcomes • Table 5.4: quality of life outcomes
	KQ28. In adults with osteoarthritis, what are the relative benefits of different patient self-management programmes i) in relation to each other or ii) versus no specific self-management programmes, with respect to symptoms, function and quality of life?	No		Yes • Table 5.1: OA site (knee and/or hip); pain outcomes • Table 5.4: quality of life outcomes
	KQ29. What is known of patient experiences of osteoarthritis and its treatments and how do patient including qualitative perceptions and beliefs influence their preference research and outcome for individual treatments?	No		No

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
BREAST CANCER				
National Collaborating Centre for Cancer[123]	KQ1. What is the role of breast MRI in the preoperative staging of patients with biopsy-proven DCIS or invasive breast cancer?	• Severity, stage, or site	For clinical questions about interventions, the PICO framework was used. This structured approach divides each question into four components: the patients (the population under study – P),the interventions (what is being done - I), the comparisons (other main treatment options – C) and the outcomes (the measures of how effective the interventions have been – O). Where appropriate, the clinical questions were refined once the evidence had been searched and, where necessary, sub-questions were generated. (*Note: For the purposes of this study, this not considered an explicit description of a priori selection*).	Yes; p 6: Patients with early, invasive breast cancer who have not yet received definitive surgery; Patients with DCIS who have received definitive surgery.
	KQ2. What is the role of pretreatment ultrasound assessment in staging the axilla?	No		Yes; p 78: Patients with early invasive breast cancer who require staging of the axilla and staging procedure planned is less than an axillary clearance.
	KQ3. What are the effective strategies to prevent and manage psychological distress in patients with early stage breast cancer?	• Severity, stage, or site		Yes; p 133: Patients with early stage breast cancer (with clinically manifest psychological distress)
	KQ4. What is the optimal tumour-free tissue margin to achieve in patients who undergo wide local excision for (DCIS)?	• Severity, stage, or site		Yes; p 206: Patients with DCIS
	KQ5. What is the role of mastectomy in patients with localised Paget's disease of the nipple?	• Severity, stage, or site		Yes; p 256: patients treated with mastectomy or breast conserving surgery for Paget's disease

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ6. In patients with invasive breast cancer or DCIS when is sentinel lymph node biopsy justified as a staging procedure?	• Severity, stage, or site		Yes; p 311: Patients with invasive breast cancer; Patients with DCIS or microinvasive Carcinoma (defined as invasive carcinoma <1mm in size)
	KQ7. What are the indications for completion axillary clearance when the axilla has been found by biopsy to contain metastasis?	• Severity, stage, or site		Yes; p 497: Patients with invasive breast cancer with histologically positive axillary nodes demonstrated by a surgical procedure i.e. SLNB or 4 node sampling
	KQ8. What is the prognostic significance of small metastatic deposits in sentinel nodes?	• Severity, stage, or site		Yes; p 549: Patients who receive SLNB as staging surgery
	KQ9. When is it appropriate to perform immediate breast reconstructive surgery?	No		Yes, p 600: Patients with breast cancer who undergo total breast reconstruction following mastectomy
	KQ10. Does progesterone receptor status add further, useful information to that of oestrogen receptor status in patients with invasive breast cancer?	• Severity, stage, or site		Yes; p 686: Patients with invasive breast cancer
	KQ11. What are the indications for adjuvant chemotherapy in patients with early invasive breast cancer?	• Severity, stage, or site		No
	KQ12. What is the optimal time interval from completion of definitive surgery to commencement of adjuvant therapy?	• Co-interventions		Yes; p 713: Patients who have received definitive surgery (including simultaneous reconstructive surgery) and who receive adjuvant therapy

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ13. In premenopausal patients with breast cancer, what are the benefits of adjuvant ovarian suppression/ablation?	• Severity, stage, or site		Yes; p 771: Premenopausal patients with early invasive breast cancer. Including: Receptor positive and Receptor negative
	KQ14. What is the role of aromatase inhibitors (AIs) as adjuvant therapy in postmenopausal women with hormone receptor positive breast cancer?.	• Severity, stage, or site		Yes, p. 840: Post menopausal patients with early invasive hormone receptor positive breast cancer who have completed definitive surgery
	KQ15. Which subgroups of post menopausal breast cancer patients should receive Aromatase Inhibitors as adjuvant therapy?	• Severity, stage, or site		Yes, p. 840: Post menopausal patients with early invasive hormone receptor positive breast cancer who have completed definitive surgery
	KQ16. Is there an indication for the use of tamoxifen after excision of pure DCIS?	• Severity, stage, or site • Co-interventions		Yes; p 964: Women who have had complete surgical excision of pure DCIS
	KQ17. What are the indications for the measurement of bone mineral density in patients with invasive breast cancer who are on adjuvant hormonal therapy?	• Co-interventions		Yes; p1118: patients who underwent breast cancer treatment
	KQ18. What are the indications for the use of bisphosphonates in patients with early breast cancer?	• Severity, stage, or site		Yes; p 1124: Patients with invasive breast cancer
	KQ19. What are the indications for radiotherapy after breast conserving surgery?	• Co-interventions		Yes; p 1116: Patients with operable invasive breast cancer who have received breast conserving surgery

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ20. When should patients with DCIS who have undergone complete excision or wide local excision (WLE) be given radiotherapy?.	• Co-interventions		Yes; p 1228: Patients with DCIS including those with microinvasive tumours who have received WLE
	KQ21. Which groups of patients should receive chest wall radiotherapy after mastectomy?	• Severity, stage, or site		Yes; p 1205: Patients treated for invasive breast cancer with mastectomy (excluding patients with DCIS)
	KQ22. What is the most effective radiotherapy dose fractionation regimen for patients receiving external beam radiotherapy after surgical excision of the breast?	No		Yes; p 1347: Patients with early invasive breast cancer • following BCS • following Mastectomy
	KQ23. What are the indications for an external beam radiotherapy boost to the site of local excision after breast conserving surgery?	• Co-interventions • Severity, stage, or site		Yes; p 1423: Patients with invasive breast cancer (not DCIS) who have received breast conserving surgery
	KQ24. What are the indications for radiotherapy to the supraclavicular fossa, internal mammary chain and axilla?	• Severity, stage, or site		Yes; p 1485: Patients with Operable invasive breast cancer who have received surgery (mastectomy or Breast conserving surgery) – need to avoid palliative surgery
	KQ25. What is the role of primary medical treatment in patients with early, invasive breast cancer?	• Severity, stage, or site		Yes; p 1565: Elderly patients: unfit for surgery or who decide not to receive surgery; patients who receive primary medical therapy with the aim of breast conserving surgery

Table C7. National Institute for Health and Clinical Excellence (NICE) (continued)

Reference	KQ2a. Key Question(s)	KQ2b. CH or subgroups addressed in KQ?	If yes, how were subgroups selected (e.g., expert opinion, literature-based, not stated explicitly, etc.) and in what section?	KQ3b. CH or subgroups addressed in analysis? If yes, how has CH been addressed?
	KQ26. For patients with inflammatory of locally advanced breast cancer who are treated with primary cytotoxic chemotherapy, what is the role of surgery and/or radiotherapy?	• Co-interventions		Yes; p 1636: Patients with inflammatory breast cancer stage III/T3-4 (locally advanced breast cancer) who have received primary chemotherapy.
	KQ27. What strategies are effective in preventing lymphodoema in patients with breast cancer?	No		Yes; p 1710: Patients with breast cancer who have received surgery, radiotherapy or no treatment
	KQ28. What strategies are effective in reducing arm and shoulder mobility problems after breast cancer surgery?	No		Yes; p 1773: Patients who undergo surgery due to breast cancer – axillary clearance, axillary RT, both
	KQ29. What treatments are effective and safe for use to treat patients with menopausal symptoms and invasive breast cancer or DCIS?	• Severity, stage, or site		Yes; p 1829: Patients with invasive breast cancer/DCIS and menopausal symptoms: which arise from treatment for invasive breast cancer ii) which arise independently of (e.g. present prior to) treatment for breast cancer. Consider subgroups with increased risk of breast cancer at an early age
	KQ30. What is the role of breast imaging modalities in the follow-up of patients with invasive breast cancer and in patients with DCIS?	• Severity, stage, or site		Yes; p 1907: Patients with invasive breast cancer or DCIS
	KQ31. What is the role of follow-up in patients who have been treated for breast cancer?	No		Yes - Patients treated for breast cancer, including those with DCIS

References

1. Bruening W, Launders J, Pinkney N, et al. Effectiveness of noninvasive diagnostic tests for breast abnormalities. Comparative Effectiveness Review No. 2. (Prepared by ECRI Evidence-based Practice Center under Contract No. 290-02-0019.) Rockville, MD: Agency for Healthcare Research and Quality. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm 2006.
2. Seidenfeld J, Samson DJ, Bonnell CJ, et al. Management of small cell lung cancer. Evidence Report/Technology Assessment No. 143. (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-02-0026.) AHRQ Publication No. 06-E016. Rockville, MD: Agency for Healthcare Research and Quality 2006.
3. Wilt TJ, Shamliyan T, Taylor B, et al. Comparative effectiveness of therapies for clinically localized prostate cancer. Comparative Effectiveness Review No. 13. (Prepared by Minnesota Evidence-based Practice Center under Contract No. 290-02-0009.) Rockville, MD: Agency for Healthcare Research and Quality. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm 2008.
4. Shekelle P, Rich M, Morton S, et al. Pharmacologic management of heart failure and left ventricular systolic dysfunction: effect in female, black, and diabetic patients, and cost-effectiveness. Evidence Report/Technology Assessment No. 82 (Prepared by the Southern California-RAND Evidence-based Practice Center under Contract No 290-97-0001). AHRQ Publication No. 03-E045. Rockville, MD: Agency for Healthcare Research and Quality. July 2003.
5. Viswanathan M, Visco AG, Hartmann K, et al. Cesarean delivery on maternal request. Evidence Report/Technology Assessment No. 133. AHRQ Pub. No. 06-E009. Rockville, MD: Agency for Healthcare Research and Quality 2006.
6. Wilt TJ ND, Kim C-B, Kane RL, Linabery A, Tacklind J, MacDonald R, Rutks I. Use of spirometry for case finding, diagnosis, and management of chronic obstructive pulmonary disease (COPD). Evidence Report/Technology Assessment No. 121 (Prepared by the Minnesota Evidence-based Practice Center under Contract No. 290-02-0009.) AHRQ Publication No. 05-E017-2. Rockville, MD. Agency for Healthcare Research and Quality. 2005.
7. Gartlehner G, Hansen RA, Thieda P, et al. Comparative effectiveness of second-generation antidepressants in the pharmacologic treatment of adult depression. Comparative Effectiveness Review No. 7. (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.) Rockville, MD: Agency for Healthcare Research and Quality. January 2007. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm. 2007.
8. Ip S, Bonis P, Tatsioni A, et al. Comparative effectiveness of management strategies for gastroesophageal reflux disease. Comparative Effectiveness Review No. 1. (Prepared by Tufts-New England Medical Center Evidence-based Practice Center under Contract No. 290-02-0022.) Rockville, MD: Agency for Healthcare Research and Quality. December 2005. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm 2005.
9. Matchar DB, McCrory DC, Orlando LA, et al. Comparative effectiveness of angiotensin-converting enzyme Inhibitors (ACEIs) and angiotensin II receptor antagonists (ARBs) for treating essential hypertension. Comparative Effectiveness Review No. 10. (Prepared by Duke Evidence-based Practice Center under Contract No. 290-02-0025.) Rockville, MD: Agency for Healthcare Research and Quality. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm. 2007.

10. Caughey AB, Sundaram V, Kaimal AJ, et al. Maternal and neonatal outcomes of elective induction of labor. Evidence Report/Technology Assessment No. 176. (Prepared by the Stanford University-UCSF Evidenced-based Practice Center under contract No. 290-02-0017.) AHRQ Publication No. 09-E005. Rockville, MD: Agency for Healthcare Research and Quality 2009.

11. Chou R, Helfand M, Peterson K, et al. Comparative effectiveness and safety of analgesics for osteoarthritis. Comparative Effectiveness Review No. 4. (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-02-0024.) Rockville, MD: Agency for Healthcare Research and Quality. September 2006. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm 2006.

12. Zhang M, Liu X, Li J, et al. Chinese medicinal herbs to treat the side-effects of chemotherapy in breast cancer patients. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

13. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. The Lancet. 2005;365(9472):1687-717.

14. Amarasena IU, Walters JAE, Wood-Baker R, et al. Platinum versus non-platinum chemotherapy regimens for small cell lung cancer. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 4.

15. Cardona Zorrilla AF, Reveiz L, Ospina Edgar G, et al. Palliative endobronchial brachytherapy for non-small cell lung cancer. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 2.

16. Dong R, Chen X, Wu T, et al. Elemene for the treatment of lung cancer. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

17. Shelley M, Wilt T, Coles B, et al. Cryotherapy for localised prostate cancer. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

18. De Conti P, Atallah Álvaro N, Arruda Homero O, et al. Intermittent versus continuous androgen suppression for prostatic cancer. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

19. Frobel A-K, Hulpke-Wette M, Schmidt Klaus G, et al. Beta-blockers for congestive heart failure in children. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2009.

20. Mathai M, Hofmeyr GJ. Abdominal surgical incisions for caesarean section. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

21. Anorlu RI, Maholwana B, Hofmeyr GJ. Methods of delivering the placenta at caesarean section. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

22. Navaneethan SD, Pansini F, Perkovic V, et al. HMG CoA reductase inhibitors (statins) for people with chronic kidney disease not requiring dialysis. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2009.

23. Roderick PJ, Willis NS, Blakeley S, et al. Correction of chronic metabolic acidosis for chronic kidney disease patients. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

24. Effing T, Monninkhof EM, van der Valk P, et al. Self-management education for patients with chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews: Reviews. 2007;Issue 4.

25. Puhan M, Scharplatz M, Troosters T, et al. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews: Reviews. 2009;Issue 1.

26. Puhan MA, Vollenweider D, Latshang T, et al. Exacerbations of chronic obstructive pulmonary disease: when are antibiotics indicated? A systematic review. Respiratory Research. 2007;8:30.

27. Yang IA, Fong K, Sim EHA, et al. Inhaled corticosteroids for stable chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

28. Nieuwenhuijsen K, Bültmann U, Neumeyer-Gromen A, et al. Interventions to improve occupational health in depressed people. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

29. Morriss R, Faizal Mohammad A, Jones Ashley P, et al. Interventions for helping people recognise early signs of recurrence in bipolar disorder. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

30. Hackett ML, Anderson CS, House A, et al. Interventions for preventing depression after stroke. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

31. Furtado Vivek A, Srihari V, Kumar A. Atypical antipsychotics for people with both schizophrenia and depression. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

32. Cipriani A, Rendell Jennifer M, Geddes J. Olanzapine in long-term treatment for bipolar disorder. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2009.

33. Liu Jian P, Yang H, Xia Y, et al. Herbal preparations for uterine fibroids. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2009.

34. Beaumont HH, Augood C, Duckitt K, et al. Danazol for heavy menstrual bleeding. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

35. Lethaby A, Augood C, Duckitt K, et al. Nonsteroidal anti-inflammatory drugs for heavy menstrual bleeding. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

36. Lethaby A, Irvine Gill A, Cameron Iain T. Cyclical progestogens for heavy menstrual bleeding. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

37. Wiysonge C, Shey U, Bradley HA, et al. Beta-blockers for hypertension. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

38. Woolf SH, Battista RN, Anderson GM, et al. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. A report by the Canadian Task Force on the Periodic Health Examination. J Clin Epidemiol. 1990;43(9):891-905.

39. Abalos E, Duley L, Steyn DW, et al. Antihypertensive drug therapy for mild to moderate hypertension during pregnancy. Cochrane Database Syst Rev. 2007(1):CD002252.

40. Zijdenbos IL, de Wit Niek J, van der Heijden GJ, et al. Psychological treatments for the management of irritable bowel syndrome. Cochrane Database of Systematic Reviews: Reviews 2009;Issue 1.

41. Webb AN, Kukuruzovic R, Catto-Smith AG, et al. Hypnotherapy for treatment of irritable bowel syndrome. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

42. Evans Beti W, Clark Wendy K, Moore David J, et al. Tegaserod for the treatment of irritable bowel syndrome and chronic constipation. Cochrane Database of Systematic Reviews: Reviews 2007;Issue 4.

43. Smyth R, Alldred SK, Markham C. Amniotomy for shortening spontaneous labour. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

44. Boulvain M, Kelly Anthony J, Irion O. Intracervical prostaglandins for induction of labour. Cochrane Database of Systematic Reviews: Reviews. 2008;Issue 1.

45. Li J, Zhang Q, Zhang M, et al. Intravenous magnesium for acute myocardial infarction. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2007.

46. Magee K, Campbell Samuel G, Moher D, et al. Heparin versus placebo for acute coronary syndromes. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

47. Wu T, Ni J, Wei J. Danshen (Chinese medicinal herb) preparations for acute myocardial infarction. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

48. Brouwer RW, van Raaij TM, Bierma-Zeinstra SMA, et al. Osteotomy for treating knee osteoarthritis. Cochrane Database of Systematic Reviews: Reviews. 2007;Issue 3.

49. Fransen M, McConnell S. Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

50. Laupattarakasem W, Laopaiboon M, Laupattarakasem P, et al. Arthroscopic debridement for knee osteoarthritis. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2008.

51. Boutin DL. Effectiveness of cognitive behavioral and supportive-expressive group therapy for women diagnosed with breast cancer: a review of the literature. Journal for Specialists in Group Work 2007;32(3):267-284.

52. van der Ploeg IM, van Esser S, van den Bosch MA, et al. Radiofrequency ablation for breast cancer: a review of the literature. European Journal of Surgical Oncology. 2007;33(6):673-677.

53. Yau G, Lock M, Rodrigues G. Systematic review of baseline low-dose CT lung cancer screening. Lung Cancer 2007;58(2):161-170.

54. Micames CG, McCrory DC, Pavey DA, et al. Endoscopic ultrasound-guided fine-needle aspiration for non-small cell lung cancer staging: a systematic review and metaanalysis. Chest 2007;131(2):539-548.

55. Coory M, Gkolia P, Yang IA, et al. Systematic review of multidisciplinary teams in the management of lung cancer. Lung Cancer 2008;60(1):14-21.

56. Tarnhuvud M, Wandel C, Willman A. Nursing interventions to improve the health of men with prostate cancer undergoing radiotherapy: a review. European Journal of Oncology Nursing 2007;11(4):328-339.

57. Lago RM, Singh PP, Nesto RW. Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials. Lancet 2007;370:1129-1136.

58. Nasr IA, Bouzamondo A, Hulot JS, et al. Prevention of atrial fibrillation onset by beta-blocker treatment in heart failure: a meta-analysis. European Heart Journal 2007;28(4):457-462.

59. Roberts F, Ryan GJ. The safety of metformin in heart failure. Annals of Pharmacotherapy 2007;41(4):642-646.

60. Nicholson WK, Wilson LM, Witkop CT, et al. Therapeutic management, delivery, and postpartum risk assessment and screening in gestational diabetes. Report 2008:96.

61. Press JZ, Klein MC, Kaczorowski J, et al. Does cesarean section reduce postpartum urinary incontinence: a systematic review. Birth 2007;34(3):228-237.

62. Balamuthusamy S, Srinivasan L, Verma M, et al. Renin angiotensin system blockade and cardiovascular outcomes in patients with chronic kidney disease and proteinuria: a meta-analysis. American Heart Journal 2008;155(5):791-805.

63. Strippoli GF, Navaneethan SD, Johnson DW, et al. Effects of statins in patients with chronic kidney disease: meta-analysis and meta-regression of randomised controlled trials. BMJ 2008;336:645-651.

64. Bradley-Drummond M, Dasenbrook EC, Pitz MW, et al. Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: a systematic review and meta-analysis. JAMA 2008;300(20):2407-2416.

65. Niesink A, Trappenburg JC, de Weert-van Oene GH, et al. Systematic review of the effects of chronic disease management on quality-of-life in people with chronic obstructive pulmonary disease. Respiratory Medicine 2007;101(11):2233-2239.

66. Quon BS, Gan WQ, Sin DD. Contemporary management of acute exacerbations of COPD: a systematic review and metaanalysis. Chest 2008;133(3):756-766.

67. Singh S, Loke YK, Furberg CD. Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. JAMA 2008;300(12):1439-50.

68. Sobieraj DM, White CM, Coleman CI. Benefits and risks of adjunctive inhaled corticosteroids in chronic obstructive pulmonary disease: a meta-analysis. Clinical Therapeutics 2008;30(8):1416-1425.

69. Usala T, Clavenna A, Zuddas A, et al. Randomised controlled trials of selective serotonin reuptake inhibitors in treating depression in children and adolescents: a systematic review and meta-analysis. European Neuropsychopharmacology 2008;18(1):62-73.

70. Cuijpers P, Brannmark JG, van Straten A. Psychological treatment of postpartum depression: a meta-analysis. Journal of Clinical Psychology 2008;64(1):103-118.

71. Chin AM. Clinical effects of reminiscence therapy in older adults: a meta-analysis of controlled trials. Hong Kong Journal of Occupational Therapy 2007;17(1):10-22.

72. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. Canadian Medical Association Journal 2008;178(3):296-305.

73. Hiyama T, Yoshihara M, Matsuo K, et al. Meta-analysis of the effects of prokinetic agents in patients with functional dyspepsia. Journal of Gastroenterology and Hepatology 2007;22(3):304-310.

74. Jin X, Li YM. Systematic review and meta-analysis from Chinese literature: the association between Helicobacter pylori eradication and improvement of functional dyspepsia. Helicobacter 2007;12(5):541-546.

75. Wang WH, Huang JQ, Zheng GF, et al. Effects of proton-pump inhibitors on functional dyspepsia: a meta-analysis of randomized placebo-controlled trials. Clinical Gastroenterology and Hepatology 2007;5(2):178-185.

76. Horvath K, Jeitler K, Siering U, et al. Long-term effects of weight-reducing interventions in hypertensive patients: systematic review and meta-analysis. Archives of Internal Medicine. 2008;168(6):571-580.

77. Connell P, Wolfe C, McKevitt C. Preventing stroke: a narrative review of community interventions for improving hypertension control in black adults. Health and Social Care in the Community. 2008;16(2):165-187.

78. Ford AC, Talley NJ, Velhuyzen van Zenten SJ, et al. Will the history and physical examination help establish that irritable bowel syndrome is causing this patient's lower gastrointestinal tract symptoms? JAMA 2008;300(15):1793-1805.

79. Ford AC, Talley NJ, Spiegel BM, et al. Effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome: systematic review and meta-analysis. BMJ 2008;337:a2313.

80. Ford AC, Chey WD, Talley NJ, et al. Yield of diagnostic tests for celiac disease in individuals with symptoms suggestive of irritable bowel syndrome. Archives of Internal Medicine 2009;169(7):651-658.

81. Rahimi R, Nikfar S, Abdollahi M. Selective serotonin reuptake inhibitors for the management of irritable bowel syndrome: a meta-analysis of randomized controlled trials. Archives of Medical Science 2008;4(1):71-76.

82. Baker WL, Couch KA. Azithromycin for the secondary prevention of coronary artery disease: a meta-analysis. American Journal of Health-System Pharmacy 2007;64(8):830-836.

83. Ioannidis JP, Katritsis DG. Percutaneous coronary intervention for late reperfusion after myocardial infarction in stable patients. American Heart Journal 2007;154(6):1065-1071.

84. Sinno MC, Khanal S, Al-Mallah MH, et al. The efficacy and safety of combination glycoprotein IIbIIIa inhibitors and reduced-dose thrombolytic therapy-facilitated percutaneous coronary intervention for ST-elevation myocardial infarction: a meta-analysis of randomized clinical trials. American Heart Journal 2007;153(4):579-586.

85. Christensen R, Bartels EM, Astrup A, et al. Symptomatic efficacy of avocado-soybean unsaponifiables (ASU) in osteoarthritis (OA) patients: a meta-analysis of randomized controlled trials. Osteoarthritis and Cartilage 2008;16(4):399-408.

86. Minns Lowe CJ, Barker KL, Dewey M, et al. Effectiveness of physiotherapy exercise after knee arthroplasty for osteoarthritis: systematic review and meta-analysis of randomised controlled trials. BMJ 2007;335:812.

87. Pisters MF, Veenhof C, van Meeteren NL, et al. Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review. Arthritis and Rheumatism 2007;57(7):1245-1253.

88. Chou R, Helfand M, Carson S. Drug class review on angiotensin converting enzyme inhibitors. Final Report. Portland, OR: Oregon Health & Science University; 2005.

89. Furmaga E, Glassman P, Rhodes S, et al. Drug class review on angiotensin II receptor antagonists. Final report. Portland, OR: Oregon Health & Science University 2006.

90. Helfand M, Peterson K, Christensen V, et al. Drug class review on beta adrenergic blockers. Final report update 4. Portland, OR: Oregon Health & Science University 2009.

91. Jonas DE, Kiser K, Shilliday BB, et al. Drug class review on controller medications for asthma. Portland, OR: Oregon Health & Science University 2008.

92. Hansen RA, Gartlehner G, Lohr KN, et al. Drug class review on inhaled corticosteroids. Final report update 1. Portland, OR: Oregon Health & Science University 2006.

93. Gartlehner G, Morgan LC, Thieda P, et al. Drug class review second generation antidepressants. Final report update 4. Portland, OR: Oregon Health & Science University 2008.

94. McDonagh MS, Carson S, Thakurta S. Drug class review on proton pump inhibitors. Final report update 5. Portland, OR: Oregon Health & Science University 2009.

95. McDonagh MS, Eden KB, Peterson K. Drug class review on calcium channel blockers. Final report. Portland, OR: Oregon Health & Science University 2005.

96. Gartlehner G, Jonas DE, Morgan LC, et al. Drug class review on constipation drugs. Final report. Portland, OR: Oregon Health & Science University 2007.

97. Dailey JH, Glassman P, McDonagh M, et al. Drug class review on newer antiplatelet agents. Final report update 1. Portland, OR: Oregon Health & Science University 2007.

98. Helfand M, Carson S, Kelley C. Drug class review on HMG-CoA reductase inhibitors (statins). Final report. Portland, OR: Oregon Health & Science University 2006.

99. Chou R, Helfand M, Peterson K, et al. Drug class review on cyclo-oxygenase (COX)-2 inhibitors and non-steroidal anti-inflammatory drugs (NSAIDs). Final report update 3. Portland, OR: Oregon Health & Science University 2006.

100. Adelaide Health Technology A. Digital mammography for breast cancer screening and diagnosis. Adelaide: Adelaide Health Technology Assessment (AHTA) 2008.

101. Korencan A, Felder-Puig R, Guba B, et al. Testing for HER2 positive breast cancer. Challenge for improvement of current conditions and practice. Wien: Ludwig Boltzmann Institut fuer Health Technology Assessment (LBIHTA) 2007.

102. Dunfield L, Severn M. Effectiveness of magnetic resonance imaging (MRI) screening for women at high risk of breast cancer. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH) 2007.

103. Adelaide Health Technology Assessment on behalf of National Horizon Scanning U. Genetic Screening for Lung Cancer utilising CT; horizon scanning prioritizing summary - volume 15. Adelaide: Adelaide Health Technology Assessment (AHTA) on behalf of National Horizon Scanning Unit (HealthPACT and MSAC) 2007.

104. Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen. Interstitial brachytherapy in localised prostate cancer. Cologne: Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2007.

105. Pearson SD, Ladapo J, Prosser L. Intensity modulated radiation therapy (IMRT) for localized prostate cancer. Boston: Institute for Clinical and Economic Review (ICER) 2007.

106. Swedish Council on Technology Assessment in Health C. Light therapy for depression, and other treatment of seasonal affective disorder. Stockholm: Swedish Council on Technology Assessment in Health Care (SBU) 2007.

107. Swedish Council on Technology Assessment in Health C. Computer-based cognitive behavioral therapy for anxiety disorders or depression - early assessment briefs (Alert). Stockholm: Swedish Council on Technology Assessment in Health Care (SBU) 2007.

108. McLoughlin DM, Mogg A, Eranti S, et al. The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis. Health Technol Assess. 2007 Jul;11(24):1-54.

109. Screening for postnatal depression within the Well Child Tamariki Ora Framework. Christchurch: Health Services Assessment Collaboration (HSAC) 2008.

110. Swedish Council on Technology Assessment in Health C. Dyspepsia and gastro-oesophageal reflux. Stockholm: Swedish Council on Technology Assessment in Health Care (SBU) 2007.

111. Hermanowski T, Kowalik E, Jakubczyk M, et al. Use of bosentan, epoprostenol, iloprost, sildenafil and treprostinil in treatment of pulmonary arterial hypertension in Poland. Budget impact analysis. Warsaw: Agency for Health Technology Assessment in Poland (AHTAPol) 2007.

112. Hermanowski T, Kowalik E, Jakubczyk M, et al. Treatment of pulmonary arterial hypertension in Poland - current practice. Survey results. Warsaw: Agency for Health Technology Assessment in Poland (AHTAPol) 2007.

113. Tran K, Ho C, Noorani HZ, et al. Overview of thiazide diuretics as first-line treatment for hypertension. [Technology overview number 34]. Ottawa: Canadian Agency for Drugs and Technologies in Health 2007.

114. Comparison of efficacy, effectiveness and safety of fibrinolysis and primary percutaneous coronary intervention (PPCI) for ST-segment elevation myocardial infarction (STEMI). Montreal: Agence d'Evaluation des Technologies et des Modes d'Intervention en Sante (AETMIS) 2008.

115. National Coordinating Centre for Health Technology Assessment. The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Home-based compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence 2007.

116. National Institute for Health and Clinical Excellence. Secondary prevention in primary and secondary care for patients following a myocardial infarction. London: National Institute for Health and Clinical Excellence (NICE) 2007.

117. Canadian Agency for Drugs and Technologies in Health (CADTH). Intra-articular hyaluronic acid (viscosupplementation) for hip osteoarthritis. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH) 2007.

118. Samson DJ, Grant MD, Ratko TA, et al. Treatment of primary and secondary osteoarthritis of the knee. Rockville: Agency for Healthcare Research and Quality (AHRQ) 2007.

119. National Collaborating Centre for Mental Health. Depression: management of depression in primary and secondary care. Rushden, Northamptsonshire: The British Psychological Society and Gaskell 2004.

120. North of England Dyspepsia Guideline Development Group. Dyspepsia: managing dyspepsia in adults in primary care. Newcastle upon Tyne (UK): Centre for Health Services Research School of Population and Health Sciences University of Newcastle upon Tyne, Report No. 112 2004.

121. Cooper A, Skinner J, Nherera L, et al. Clinical guidelines and evidence review for post myocardial infarction: secondary prevention in primary and secondary care for patients following a myocardial infarction. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners 2007.

122. National Collaborating Centre for Chronic Conditions. Osteoarthritis: national clinical guideline for care and management in adults. London: Royal College of Physicians 2008.

123. National Collaborating Centre for Cancer. Early and locally advanced breast cancer: diagnosis and treatment. *NICE Clinical Guideline*. London (UK): National Institute for Health and Clinical Excellence (NICE) 2009:37.

Table C8.

Paper	Does it distinguish clinical from methodologic heterogeneity?	Does it address heterogeneity within studies?	Does it address heterogeneity among studies?	Does it give guidance for identifying measures of clinical heterogeneity?
Adams, 1998[1]	Yes	Yes	Yes	Yes
Anello and Fleiss, 1995 [2]	No	No	Yes	No
Bailar, 1995[3]	No	No	Yes	No
Bailey, 1987[4]	No	No	Yes	No
Barza, Trinkalinos, and Lau, 2009[5]	Yes	Yes	Yes	No
Baujat et al., 2002[6]	No	Yes	Yes	No
Bellinger, 2000[7]	Yes	Yes	Yes	Yes
Berlin, 1995[8]	Yes	Yes	Yes	No
Berlin and Colditz, 1999[9]	Yes	Yes	Yes	No
Berlin et al., 2002[10]	Yes	Yes	Yes	No
Berry, 1990[11]	No	Yes	No	No
Bigger , 2003[12]	Yes	Yes	Yes	Yes
Blettner et al., 1999[13]	No	Yes	Yes	No
Brandand Kragt, 1992[14]	No	Yes	Yes	No
Brookes et al., 2004[15]	No	Yes	Yes	No
Brookes et al., 2001[16]	No	Yes	No	No
Chalmers, 1991[17]	No	No	Yes	No
Chalmers and Lau, 1996[18]	No	No	Yes	No
Colditz, Burdick, and Mosteller, 1995[19]	Yes	No	Yes	Yes
Cook, Sackett, and Spitzer, 1995 [20]	No	No	Yes	No
Counsell, 1997[21]	No	No	Yes	No
Delgado-Rodriguez, 2006[22]	No	No	Yes	No
Dickersin and Berlin, 1992[23]	Yes	Yes	Yes	No
Donner and Klar, 2002[24]	Yes	Yes	Yes	No
Engels et al., 2000[25]	No	No	Yes	No
Eysenck, 1994[26]	Yes	No	Yes	No
Feinstein, 1998[27]	No	Yes	No	Yes
Feinstein, 2996[28]	No	Yes	Yes	No
Fletcher, 2002[29]	Yes	No	Yes	No
Freemantle, Mason, and Eccles, 1999[30]	Yes	Yes	Yes	No
Gerritsen et al., 2002[31]	Yes	No	Yes	No
Glasziou and Sanders, 2002[32]	Yes	Yes	Yes	No

Table C8. (continued)

Paper	Does it distinguish clinical from methodologic heterogeneity?	Does it address heterogeneity within studies?	Does it address heterogeneity among studies?	Does it give guidance for identifying measures of clinical heterogeneity?
Greenland and O'Rourke, 2008[33]	Yes	Yes	Yes	No
Hahn et al., 2000[34]	No	Yes	Yes	Yes
Hardy and Thompson, 1998[35]	No	Yes	Yes	Yes
Higgins et al., 2002[36]	Yes	Yes	Yes	Yes
Higgins et al., 2003[37]	Yes	No	Yes	No
Horowitz, 2987[38]	Yes	Yes	Yes	Yes
Ioannidis, Patsopoulos, and Evanvelou, 2007[39]	Yes	No	Yes	No
Ioannidis and Lau, 1997[40]	No	Yes	Yes	No
Ioannidis and Lau, 1998[41]	Yes	Yes	Yes	No
Ioannidis, Trikalinos, and Zintzaras, 2006[42]	Yes	Yes	Yes	No
Koopman et al., 2007[43]	Yes	Yes	Yes	No
Kravitz, Duan, and Braslow, 2004[44]	Yes	Yes	Yes	No
Lambert et al., 2002[45]	Yes	Yes	Yes	No
Lau, Ioannidis, and Schmid, 1997[46]	Yes	Yes	Yes	No
Lau, Ioannidis, and Schmid, 1998[47]	Yes	Yes	Yes	No
Lecky, Little, and Brennan, 1996[48]	Yes	Yes	Yes	No
Liberati, 1995[49]	Yes	No	Yes	No
MacArthur, Foran, and Bailar, 1995[50]	Yes	No	Yes	No
Maguire et al., 2008[51]	Yes	Yes	Yes	No
Messori, 1997[52]	No	No	Yes	No
Moher et al., 2000[53]	Yes	No	Yes	No
Moher, Jadad, and Klassen, 1998[54]	Yes	Yes	Yes	No
Mosteller and Colditz, 1996[55]	No	No	Yes	No
Oxman and Guyatt, 1992[56]	Yes	Yes	Yes	No
Patsopoulos, Evangelou, and Ioannidis, 2008[57]	No	No	Yes	No
Petitti, 1997[58]	No	No	Yes	No
Petitti, 2001[59]	Yes	No	Yes	No
Reade et al., 2008[60]	Yes	Yes	Yes	No
Riley et al., 2008[61]	No	Yes	Yes	No
Rothwell, 2005[62]	Yes	Yes	Yes	Yes
Santaguida, Helfand, and Raina, 2005[63]	Yes	Yes	Yes	Yes
Sauerland and Seiler, 2005[64]	Yes	Yes	Yes	No
Schmid et al., 1998[65]	Yes	Yes	Yes	No
Schmid et al., 2004[66]	Yes	Yes	Yes	No

Table C8. (continued)

Paper	Does it distinguish clinical from methodologic heterogeneity?	Does it address heterogeneity within studies?	Does it address heterogeneity among studies?	Does it give guidance for identifying measures of clinical heterogeneity?
Schmid, 1999[67]	Yes	Yes	Yes	Yes
Simmonds et al., 2005[68]	Yes	Yes	Yes	No
Smith and Egger, 2001[69]	Yes	Yes	Yes	Yes
Smith, Williamson, and Marson, 2005[70]	Yes	Yes	Yes	No
Song et al., 2001[71]	Yes	Yes	Yes	No
Sutton and Higgins, 2008[72]	Yes	Yes	Yes	No
Thacker, Peterson, and Stroup, 1996[73]	No	Yes	Yes	No
Thompson, 1994[74]	Yes	Yes	Yes	No
Thompson and Higgins, 2002[75]	Yes	Yes	Yes	No
Thompson and Higgins, 2005[76]	No	Yes	Yes	Yes
Thompson and Pocock, 1991[77]	Yes	No	Yes	No
Thompson and Sharp, 1999[78]	No	Yes	Yes	No
Thompson, Turner, and Warn, 2001[79]	No	Yes	Yes	No
Villar, Carroli, and Belizan, 1995[80]	No	No	Yes	No
Weed, 2000[81]	Yes	No	Yes	No
Winegardner et al., 2007[82]	Yes	Yes	Yes	No
Xu et al., 2008[83]	Yes	Yes	Yes	No

References

1. Adams KFJ. Post hoc subgroup analysis and the truth of a clinical trial [editorial]. American Heart Journal 1998 Nov;136(5):753-758.
2. Anello C, Fleiss JL. Exploratory or analytic meta-analysis: should we distinguish between them? J Clin Epidemiol 1995 Jan;48(1):109-116; discussion 17-18.
3. Bailar 3rd JC. The practice of meta-analysis. J Clin Epidemiol. 1995 Jan;48(1):149-157.
4. Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? Stat Med 1987 Apr-May;6(3):351-360.
5. Barza M, Trikalinos TA, Lau J. Statistical considerations in meta-analysis. Infectious Disease Clinics of North America 2009 Jun;23(2):195-210, Table of Contents.
6. Baujat B, Mahe C, Pignon JP, et al. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652.
7. Bellinger DC. Effect modification in epidemiologic studies of low-level neurotoxicant exposures and health outcomes 2000:133-140.
8. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. Am J Epidemiol 1995 Aug;142(4):383-387.
9. Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA 1999 Mar 3;281(9):830-834.
10. Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002 Feb 15;21(3):371-387.
11. Berry DA. Subgroup analyses [letter]. Biometrics 1990 Dec;46(4):1227-1230.
12. Bigger JT. Issues in subgroup analyses and meta-analyses of clinical trials. J Cardiovasc Electrophysiol 2003 Sep;14(9 Suppl):S6-S8.
13. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. International Journal of Epidemiology 1999 Feb;28(1):1-9.
14. Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. Stat Med 1992 Dec;11(16):2077-2082.
15. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. Journal of Clinical Epidemiology 2004 Mar;57(3):229-236.
16. Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technology Assessment 2001;5(33):1-56.
17. Chalmers TC. Problems induced by meta-analyses. Stat Med. 1991 Jun;10(6):971-979; discussion 9-80.
18. Chalmers TC, Lau J. Changes in clinical trials mandated by the advent of meta-analysis. Stat Med 1996 Jun 30;15(12):1263-1268; discussion 9-72.
19. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. Am J Epidemiol 1995 Aug 15;142(4):371-382.
20. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995 Jan;48(1):167-171.
21. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Annals of Internal Medicine 1997 Sep;127(5):380-387.
22. Delgado-Rodriguez M. Systematic reviews of meta-analyses: applications and limitations. Journal of Epidemiology and Community Health 2006 Feb;60(2):90-92.
23. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science Epidemiologic Reviews 1992;14:154-176.

24. Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. Stat Med 2002 Oct 15;21(19):2971-2980.
25. Engels EA, Schmid CH, Terrin N, et al. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med 2000 Jul 15;19(13):1707-1728.
26. Eysenck HJ. Meta-analysis and its problems. BMJ 1994 Sep 24;309(6957):789-792.
27. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy [comment]. Journal of Clinical Epidemiology 1998 Apr;51(4):297-299.
28. Feinstein AR. Meta-analysis and meta-analytic monitoring of clinical trials. Stat Med 1996 Jun 30;15(12):1273-1280; discussion 81-83.
29. Fletcher RH. Evaluation of interventions. J Clin Epidemiol 2002 Dec;55(12):1183-1190.
30. Freemantle N, Mason J, Eccles M. Deriving treatment recommendations from evidence within randomized trials - The role and limitation of meta-analysis. International Journal of Technology Assessment in Health Care 1999 Spr;15(2):304-315.
31. Gerritsen AAM, de Vet HCW, Scholten R, et al. Enabling meta-analysis in systematic reviews on carpal tunnel syndrome. Journal of Hand Surgery-American Volume 2002 Sep;27A(5):828-832.
32. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. Stat Med 2002 Jun 15;21(11):1503-1511.
33. Greenland S, O'Rourke K. Meta-Analysis. In: Rothman KJ, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins 2008:652-682.
34. Hahn S, Williamson PR, Hutton JL, et al. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med 2000 Dec 30;19(24):3325-3336.
35. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Statistics in Medicine 1998 Apr;17(8):841-856.
36. Higgins J, Thompson S, Deeks J, et al. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. Journal of Health Services Research & Policy 2002 Jan;7(1):51-61.
37. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560.
38. Horwitz RI. Complexity and contradiction in clinical trial research. The American Journal of Medicine 1987 Mar;82(3):498-510.
39. Ioannidis JPA, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007;335(7626):914-916.
40. Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. J Clin Epidemiol 1997 Oct;50(10):1089-1098.
41. Ioannidis JP, Lau J. Uncontrolled pearls, controlled evidence, meta-analysis and the individual patient. J Clin Epidemiol 1998 Aug;51(8):709-711.
42. Ioannidis JP, Trikalinos TA, Zintzaras E. Extreme between-study homogeneity in meta-analyses could offer useful insights. J Clin Epidemiol 2006 Oct;59(10):1023-1032.
43. Koopman L, van der Heijden GJ, Glasziou PP, et al. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. J Clin Epidemiol 2007 Oct;60(10):1002-1009.
44. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 2004;82(4):661-687.
45. Lambert PC, Sutton AJ, Abrams KR, et al. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002 Jan;55(1):86-94.
46. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997 Nov 1;127(9):820-826.
47. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. Lancet 1998 Jan 10;351(9096):123-127.
48. Lecky FE, Little RA, Brennan P. The use and misuse of meta-analysis. Journal of Accident & Emergency Medicine 1996 Nov;13(6):373-378.

49. Liberati A. "Meta-analysis: statistical alchemy for the 21st century": discussion. A plea for a more balanced view of meta-analysis and systematic overviews of the effect of health care interventions. J Clin Epidemiol 1995 Jan;48(1):81-86.

50. Macarthur C, Foran PJ, Bailar 3rd JC. Qualitative assessment of studies included in a meta-analysis: DES and the risk of pregnancy loss. J Clin Epidemiol 1995 Jun;48(6):739-747.

51. Maguire MJ, Hemming K, Hutton JL, et al. Overwhelming heterogeneity in systematic reviews of observational anti-epileptic studies. Epilepsy Research 2008 Aug;80(2-3):201-212.

52. Messori A. Current controversies in the application of meta-analysis (with special reference to oncological treatments). Pharmacy World & Science 1997 Jun;19(3):152-158.

53. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Onkologie 2000 Dec;23(6):597-602.

54. Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews-III. How did the authors synthesize the data and make their conclusions? Archives of Pediatrics & Adolescent Medicine 1998 Sep;152(9):915-920.

55. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). Annual Review of Public Health 1996;17:1-23.

56. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992 Jan 1;116(1):78-84.

57. Patsopoulos NA, Evangelou E, Ioannidis JP. Heterogeneous views on heterogeneity. International Journal of Epidemiology 2008 Oct 21.

58. Petitti DB. Meta-analysis and endocrinology. Endocrinology and Metabolism Clinics of North America 1997 Mar;26(1):31-37.

59. Petitti DB. Approaches to heterogeneity in meta-analysis. Statistics in Medicine 2001 Dec;20(23):3625-3633.

60. Reade MC, Delaney A, Bailey MJ, et al. Bench-to-bedside review: avoiding pitfalls in critical care meta-analysis--funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy. Critical Care 2008;12(4):220.

61. Riley RD, Dodd SR, Craig JV, et al. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Statistics in Medicine 2008 Dec;27(29):6111-6136.

62. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005 Jan 1-7;365(9453):82-93.

63. Santaguida PL, Helfand M, Raina P. Challenges in systematic reviews that evaluate drug efficacy or effectiveness. Annals of Internal Medicine 2005 Jun 21;142(12 Pt 2):1066-1072.

64. Sauerland S, Seiler CM. Role of systematic reviews and meta-analysis in evidence-based medicine. World Journal of Surgery 2005 May;29(5):582-587.

65. Schmid CH, Lau J, McIntosh MW, et al. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 1998 Sep 15;17(17):1923-1942.

66. Schmid CH, Stark PC, Berlin JA, et al. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. Journal of Clinical Epidemiology 2004 Jul;57(7):683-697.

67. Schmid CH. Exploring heterogeneity in randomized trials via meta-analysis. Drug Information Journal. 1999;33:211-224.

68. Simmonds MC, Higgins JP, Stewart LA, et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clinical Trials 2005;2(3):209-217.

69. Smith GD, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials [book chapter]. In: Egger M, Smith GD, Altman DG, eds. Systematic Review in Health Care: Meta-analysis in Context. 2nd ed. London, England: BMJ Publishing Group 2001:143-156.

70. Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. Journal of Evaluation in Clinical Practice 2005 Oct;11(5):468-478.

71. Song FJ, Sheldon TA, Sutton AJ, et al. Methods for exploring heterogeneity in meta-analysis. Evaluation & the Health Professions 2001 Jun;24(2):126-151.

72. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Stat Med 2008 Feb 28;27(5):625-650.

73. Thacker SB, Peterson HB, Stroup DF. Metaanalysis for the obstetrician-gynecologist. American Journal of Obstetrics and Gynecology 1996 May;174(5):1403-1407.

74. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994 Nov 19;309(6965):1351-1355.

75. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002 Jun 15;21(11):1559-1573.

76. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005 Jan 22-28;365(9456):341-346.

77. Thompson SG, Pocock SJ. Can meta-analyses be trusted? Lancet 1991 Nov 2;338(8775):1127-1130.

78. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. Statistics in Medicine 1999 Oct;18(20):2693-2708.

79. Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk differences. Statistical Methods in Medical Research 2001 Dec;10(6):375-392.

80. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. Lancet 1995 Mar 25;345(8952):772-776.

81. Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. International Journal of Epidemiology 2000 Jun;29(3):387-390.

82. Winegardner ML, Reaume KT, Dabaja GS, et al. Practical application of nonrandomized research to patient care: A case study of nesiritide. Pharmacotherapy 2007 Jan;27(1):143-151.

83. Xu HR, Platt RW, Luo ZC, et al. Exploring heterogeneity in meta-analyses: needs, resources and challenges 2008; 2008. p. 18-28.

Appendix D. Acknowledgments

This study was supported by Contract 290-2007-10056 I from the Agency for Healthcare Research and Quality (AHRQ), Task No.7, Work Assignment 9 and Task No. 2, Work Assignment 4 under Contract 290-2007-10056I II. We acknowledge the continuing support of Beth Collins Sharp, Ph.D., R.N. Director of the AHRQ Evidence-based Practice Center (EPC) Program and Stephanie Chang, M.D., MPH the AHRQ Task Order Officer for this project.

The investigators appreciate the guidance of Mark Helfand, M.D., M.P.H. from the Oregon Health Sciences Center. Also from Oregon Health Sciences Center, we thank Howard Balshem, M.S. and James Wallace, B.A. for their assistance with obtaining the peer review documents we analyzed for this report and Rose Campbell, M.L.I.S., M.S. for providing additional methods references for us to review. We express our gratitude to Loraine Monroe, EPC publication specialist at RTI International.

Peer Review

Peer reviewers read and provided feedback on a draft version of the report. We revised the report as appropriate in response to their suggestions.

Peer Reviewers	Affiliation
David Atkins, M.D., M.P.H.	VA Quality Enhancement Research Initiative (QUERI) Washington, DC 20420
Jesse Berlin, Sc.D.	Johnson & Johnson 1125 Trenton-Harbourton Road; PO Box 200; MS K-304; Titusville, NJ 08560
Jean-Paul Gagnon, Ph.D.	Sanofi-Aventis 55 Corporate Drive Bridgewater, NJ 08807
Sheldon Greenfield, M..	University of California-Irvine School of Medicine 111 Academy Suite 220 Irvine, CA 92697
Richard L. Kravitz, M.D., M.S.P.H.	UC Davis Division of General Medicine 4150 V. Street, Suite 2400 PSSB Sacramento, CA 95817
Noel Weiss, M.D., Dr.P.H.	University of Washington Box 357236 Seattle, WA 98195