Research and Theory for Nursing Practice: An International Journal, Vol. 25, No. 3, 2011

Comparative Effectiveness Research: Using Systematic Reviews and Meta-Analyses to Synthesize Empirical Evidence

Maria C. Katapodi, PhD, MSc, BSN

Laurel L. Northouse, PhD, RN, FAAN

University of Michigan School of Nursing, Ann Arbor

The increased demand for evidence-based health care practices calls for comparative effectiveness research (CER), namely the generation and synthesis of research evidence to compare the benefits and harms of alternative methods of care. A significant contribution of CER is the systematic identification and synthesis of available research studies on a specific topic. The purpose of this article is to provide an overview of methodological issues pertaining to systematic reviews and meta-analyses to be used by investigators with the purpose of conducting CER. A systematic review or meta-analysis is guided by a research protocol, which includes (a) the research question, (b) inclusion and exclusion criteria with respect to the target population and studies, (c) guidelines for obtaining relevant studies, (d) methods for data extraction and coding, (e) methods for data synthesis, and (f) guidelines for reporting results and assessing for bias. This article presents an algorithm for generating evidence-based knowledge by systematically identifying, retrieving, and synthesizing large bodies of research studies. Recommendations for evaluating the strength of evidence, interpreting findings, and discussing clinical applicability are offered.

Keywords: evidence-based practice; research

The demand for health care practices that are evidence-based, standardized, consistent, and promote optimal outcomes within the constraints of limited resources is growing. Today's health care technologies provide a wide range of interventions, tests, and strategies for prevention, risk management, and treatment of many diseases. Staying current with new evidence can be overwhelming for providers. Patient advocates suggest that patients become informed consumers within the health care system and actively choose a course of action that is in line

http://dx.doi.org/10.1891/1541-6577.25.3.191

with their own values. However, the information necessary to inform health care decisions is often incomplete or unavailable. As a result, many of the interventions and treatments used today are delivered without clear evidence of their effectiveness.

The Institute of Medicine (IOM) has called for a national research initiative to support better decision making in health care (Sox & Greenfield, 2009). This research, known as *comparative effectiveness research* (CER), aims to identify what practices work, for which patients, and under what circumstances. CER is defined as ". . . *the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care*" (IOM, 2009, p. 2). This definition emphasizes the need for direct comparisons of effective interventions to identify the care, which is most appropriate for individual patients and families. CER answers two distinct questions: (a) "Is Intervention A better than Intervention B for an individual patient?" and (b) "What are the clinical characteristics of patients who are more likely to benefit from a specific intervention?" The answers to these questions can help consumers, health care providers, purchasers, and policy makers make informed decisions that will improve health care at both the individual and population level.

Realizing the full potential of CER depends on accessing, evaluating, and synthesizing large bodies of heterogeneous studies (Sox et al., 2010). Systematic reviews and meta-analyses are based on scientific strategies that reduce bias to assemble, appraise, and synthesize relevant studies on a specific topic. Meta-analysis is a subsequent step of a systematic review that can be applied when appropriate; it is a statistical technique that pools data from many individual studies and provides a combined estimate of efficacy of a particular treatment from the overall evaluation of these studies (Lipsey & Wilson, 2001).

The purpose of this article is to provide an algorithm for the conduct of systematic reviews and meta-analyses. This algorithm can be used by investigators who seek to improve health care decision making by examining and synthesizing available evidence of alternative methods of delivering care. This article concludes with a discussion of how systematic reviews and meta-analyses promote the goals of CER and provides suggestions for future research.

METHODOLOGY FOR CONDUCTING SYSTEMATIC REVIEWS AND META-ANALYSES

Systematic reviews are often confused with general literature reviews. Literature reviews are used to describe the content on a particular topic but seldom seek all available evidence in a systematic manner, specify the search process, and state how data were extracted from studies. In contrast, systematic reviews and meta-analyses follow a clearly prescribed research protocol and describe in detail the methods used to obtain, examine, and interpret the data from available studies. In some cases, systematic reviews and meta-analyses have reported incomplete methodological information; thus, it was not possible to determine the quality of the available evidence. Consequently, the Preferred Reporting Items for Systematic Reviews

and Meta-Analyses (PRISMA) statement was developed by a group of researchers, clinicians, methodologists, health professionals, and consumers to improve the quality of these reports. The PRISMA statement is a checklist of items for authors to follow to provide more complete and transparent reports of systematic reviews and meta-analyses that evaluate health care interventions (Liberati et al., 2009; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). According to the PRISMA guidelines, any systematic review or meta-analysis should have a research protocol to guide the research process. Research protocols include the (a) research question, (b) inclusion and exclusion criteria with respect to target population and studies, (c) guidelines for obtaining relevant studies, (d) methods for data extraction and coding, (e) methods for data synthesis, and (f) guidelines for reporting results and assessing for bias. The reporting document evaluates the strength of evidence, interprets findings, discusses their clinical applicability, and examines limitations of the study. Components of a research protocol are discussed in more detail.

COMPONENTS OF A RESEARCH PROTOCOL

Formulating the Research Question

The first step of a systematic review or a meta-analysis is to determine the research question. The research question needs to be specific and narrow in focus to help guide strategies for locating, selecting, and appraising the relevance and validity of studies. A well-formulated research question specifies the key dependent variables, the major independent variables, the interventions or exposures of interest, and the population(s) being investigated (Lipsey & Wilson, 2001). Research questions may examine the *efficacy of treatments* (e.g., whether topical nonsteroidal anti-inflammatory drugs [NSAIDs] are efficacious in treating osteoarthritis; Lin, Zhang, Jones, & Doherty, 2004), *identify subgroups* of patients more likely to benefit from a treatment (e.g., whether nicotine replacement therapy is more effective for men than women; Cepeda-Benito, Reynoso, & Erath, 2004), and *identify outcomes* of specific types of interventions (e.g., quality-of-life outcomes associated with physical activity interventions; Conn, Hafdahl, & Brown, 2009).

Inclusion and Exclusion Criteria

Research protocols specify the inclusion and exclusion criteria for the *populations* and *types of studies* to be used in the review or meta-analysis.

Population-related criteria refer to "condition" of interest (e.g., coronary artery disease), "patients" of interest (e.g., subgroups of patients according to age, gender, race), and "setting" of interest (e.g., inpatient, community). Any criteria or restrictions related to these areas are based on empirical or clinical evidence. For example, Clark, Haykowsky et al. (2010) examined the effects of home-based versus hospital-based secondary prevention programs for the management of coronary artery disease. The authors justified the need to examine home-based as well as hospital-based programs by presenting empirical evidence indicating that patients' access

to hospital-based programs was limited, especially among those patients with a greater need for risk reduction, whereas home-based programs increased access among vulnerable populations.

Study-related criteria refer to research design and publication type:

Research design: Deciding on the inclusion and exclusion criteria for types of studies is a challenging task because findings can be influenced by a range of methodological characteristics. Systematic reviews can synthesize data both from observational studies and from experimental trials. However, meta-analyses *cannot* combine data from studies with different designs (i.e., observational and experimental) to provide a pooled estimate of efficacy. For example, Conn, Hafdahl, Cooper, Brown, and Lusk (2009) examined the effects of physical activity interventions delivered in workplaces. Although their meta-analyses included both observational and experimental studies, they conducted separate analyses for observational studies (e.g., descriptive and correlational) and for experimental trials with two-group comparisons. Similarly, a meta-analysis that examined the efficacy of tai chi exercise on aerobic capacity reported that effect sizes from experimental trials showed a small and insignificant effect of tai chi on aerobic capacity, which was not consistent with effect sizes obtained from observational studies (Taylor-Piliae & Froelicher, 2004).

Despite the known methodological limitations of observational studies (i.e., cannot establish causality and are susceptible to bias of small samples), meta-analysis of observational data can provide useful information. The pooled effect size obtained from observational data represents the overall strength of association between predictors and the outcome of interest (Borenstein, Hedges, Higgins, & Rothstein, 2009; Lipsey & Wilson 2001). For example, two meta-analyses of observational studies reported that perceived breast cancer risk had only a small effect on whether or not women obtained routine mammograms (Katapodi, Lee, Facione, & Dodd, 2004; McCaul, Branstetter, Schroeder, & Glasgow, 1996). On the other hand, even though experimental trials use similar designs and are considered the "gold standard" of research design, they can also have subtle methodological differences. Experimental trials may vary on the type of control groups used, blinding of study participants and allocation concealment, exclusions after randomization, and characteristics of the evaluation method. Even these subtle differences between study designs have been found to produce large differences in treatment effects (Schulz, Chalmers, Hayes, & Altman, 1995; Sellers, Crawford, Bullock, & McKinlay, 1997).

Publication type: This term refers to whether peer-reviewed journal articles, books, dissertations, technical reports, conference proceedings, studies from different countries and languages, or unpublished manuscripts will be included in the systematic review or meta-analysis. Some researchers limit their reports to only peer-reviewed articles, which are easier to locate compared with conference proceedings and technical reports.

Using broader or more inclusive study criteria provide a fuller representation of research conducted on a specific topic. It also provides the opportunity to examine synergistic effects between interventions and sample characteristics. A potential problem to using broader methodological criteria is the inclusion of studies that are

not as relevant to the topic being examined, which may lead to erroneous results. However, too strict criteria may exclude smaller studies and lead to results that are less generalizable (Lipsey & Wilson, 2001). For example, in the meta-analysis that examined the effects of physical activity interventions delivered in workplaces, Conn, Hafdahl, Cooper, Brown, and colleagues (2009) increased the generalizability of their findings by including studies with different types of interventions, diverse samples, studies with smaller samples, and unpublished reports.

Obtaining Relevant Studies

Identifying and retrieving relevant studies is complex and the most time-consuming step of a systematic review or meta-analysis. A close consultation with a reference librarian is highly recommended. The aim of the search is to generate a comprehensive list of published and unpublished studies that are suitable to answer the research question. A comprehensive search process has to be unbiased, systematic, and yield reproducible results (Borenstein et al., 2009; Lipsey & Wilson, 2001).

The first step in organizing the search process is to develop a system to record the list of available sources, the keywords and other criteria used to identify studies, the source of each citation, the type of publication, and the date when each citation was identified and retrieved (Lipsey & Wilson, 2001). Systematic reviews and meta-analyses search multiple sources (e.g., MEDLINE, PsycINFO) because it is highly unlikely that one source alone will identify all potentially eligible studies. The search process is complimented with hand searches of reference lists from review articles and specialty journals (e.g., *Heart & Lung*, *Cancer Nursing*) that may publish studies relevant to the topic (Dickersin, Scherer, & Lefebvre, 1994; McManus et al., 1998).

To effectively locate a high proportion of eligible studies, the search is based on a set of keywords that broadly cover the relevant domain (Lipsey & Wilson, 2001). Keywords are used in combinations with conjunctive and disjunctive commands in the computer search. The search process is iterative, meaning that the search criteria and the keywords used may be revised based on the studies identified. If a change is made to the search criteria and the keywords used, the change must be applied retroactively to the databases examined prior to the change (Lipsey & Wilson). Using reference management software (e.g., EndNote, RefWorks) is highly recommended for systematic management of retrieved studies.

Any search will generate a proportion of true candidate studies and many irrelevant studies. The separation of relevant from irrelevant studies is done manually and involves multiple steps. The first step is to identify and exclude duplicate study titles identified from all search sources. Potentially relevant titles should be provisionally included for consideration based on abstracts and full text articles. Duplicate publications, or publications from the same research study, are difficult to identify and are a source of bias for systematic reviews and meta-analyses (Liberati et al., 2009). Studies should be excluded only after careful consideration; once a study is considered to be irrelevant, it is highly unlikely that investigators will revise this decision and useful information may be lost (Lipsey & Wilson, 2001).

The search process concludes with a flowchart, detailing the studies included and excluded from the review and the reasons for exclusions. For example, Clark, Smith, Taylor, and Campbell (2010) conducted a systematic review and meta-analysis of nurse-led interventions designed to control blood pressure in hypertensive patients. They used a flowchart (see Figure 1) to describe the process of identifying their final sample of 33 studies included in their narrative systematic review and 32 studies included in their quantitative meta-analysis.

Methods for Data Extraction and Coding

After an agreement has been reached about which studies to include in the review, data from the selected studies are extracted and documented in customized tables following specific coding mechanisms. Coding tables could be reported as an appendix to the systematic review or meta-analysis (Moher et al., 2009). The data collected include information about the authors, publication dates, study designs, samples, characteristics of interventions delivered, the settings, the instruments used to measure the independent and dependent variables, and the study outcomes. Tables 1, 2, and 3 provide examples of tables for data extraction used in a meta-analysis of intervention studies with caregivers of patients with cancer (Northouse, Katapodi, Song, Zhang, & Mood, 2010).

There are two distinct parts in the coding protocol: one part encodes information about study characteristics and the other part encodes information about the findings of the study (Borenstein et al., 2009; Lipsey & Wilson, 2001). Study characteristics represent the *independent variables* of a systematic review or a meta-analysis, such as the types of interventions used in the study, sample characteristics, or instruments used. The *dependent variables* of the systematic review or meta-analysis are the study outcomes of the primary studies. Thus, the coding protocol should be developed in modules: the first module to code information that applies to the study characteristics and the second module to code information for outcomes and effect sizes (Lipsey & Wilson, 2001).

Data extraction involves collection and scrutiny of detailed raw data. This task is undertaken by two or more investigators with expertise on the research question. In this phase, it is not uncommon for investigators to contact primary authors and request additional information that was not reported or unclearly reported in the publication of the study. It is important to report what additional information was sought from primary investigators and whether or not it was obtained (Liberati et al., 2009). For example, Papadakis and colleagues (2010) examined the delivery of smoking cessation treatments in primary care. Despite efforts to obtain additional data from primary authors, they excluded eight eligible studies because of insufficient data in the study report.

Methods for Data Synthesis

Data can be synthesized in the form of a systematic review or meta-analysis. Systematic reviews provide a narrative synthesis of available studies. They have a broad scope and can synthesize evidence from qualitative and nonexperimental

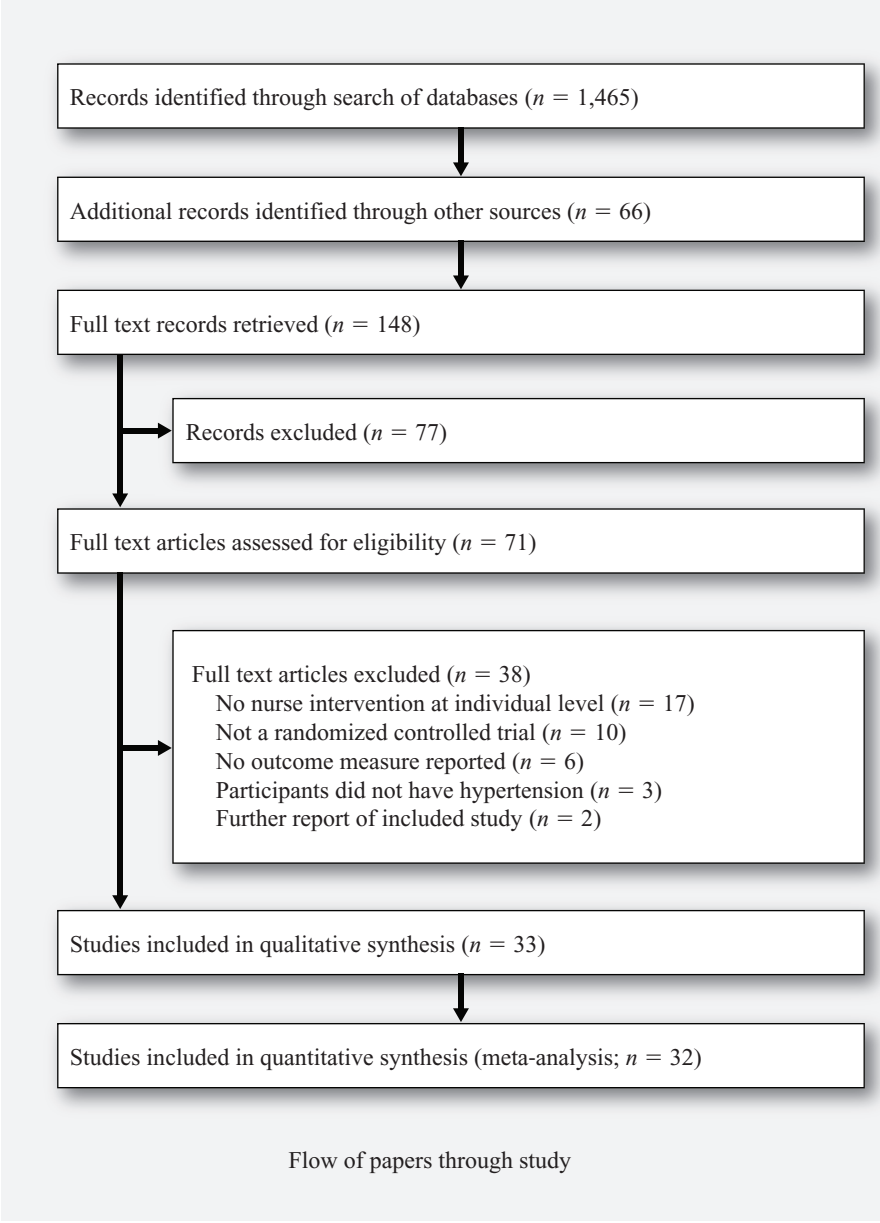

Figure 1. Example of study selection process. Reproduced from "Nurse Led Interventions to Improve Control of Blood Pressure in People With Hypertension: Systematic Review and Meta-Analysis," by C. E. Clark, L. F. Smith, R. S. Taylor, and J. L. Campbell, 2010, *British Medical Journal*, 341, p. c3995 with permission from BMJ Publishing Group Ltd.

TABLE 1. Example of Coding for Characteristics of Primary Studies

Author/ Year	Aims	Design	Setting	Sample	Outcomes	Measures	Findings
Northouse et al. (2007)	Improve: • Appraisal • Coping • Symptom distress • QoL in patients with prostate cancer and caregivers	RCT; longitudinal assessments at 4, 8, and 12 months	Patient's home	263 patient–caregiver dyads	• Illness appraisals • Coping • Symptom distress • QoL	• Appraisal of Caregiving Scale • Mishel Uncertainty in Illness Scale • Beck Hopelessness Scale • Brief COPE • Lewis Cancer Self-Efficacy Scale • Lewis Mutuality and Interpersonal Sensitivity Scale • Omega Screening Questionnaire—Symptom • MOS SF-12 • FACT-G	Four-month follow-up: • Intervention patients: - *Less* uncertainty - *Better* communication • Intervention caregivers: - *Higher* QoL and self-efficacy - *Better* communication - *Less* negative appraisals of caregiving, uncertainty, hopelessness, and symptom distress

Note. QoL = quality of life; RCT = randomized controlled trial; COPE = coping orientation for problem experiences; MOS SF-12 = medical outcomes study 12-item short form; FACT-G = functional assessment of cancer therapy—general.

TABLE 2. Example of Coding for Intervention Characteristics

Study/Primary Aim	Intervention Type/Contacts—Number, Length, Location	Duration of Protocol/Intervener	Patient Caregiving	Marital/Family Care	Caregiver Self-Care	Theoretical Framework/Fidelity
Northouse et al. (2007) • Improve appraisal, coping, symptom distress, and QoL in patients with prostate cancer and their caregivers	• Skills training • Face-to-face and telephone • Included patient and caregiver • Additional printed materials • Three joint home visits (90 minutes each) alternating with two joint telephone calls (30 minutes each)	• Thirteen weeks • Master's RNs	• Assist patient with managing care needs • Provide support • Communicate with health care providers	• Open communication • Mutual support • Healthy lifestyle behaviors • Maintain optimism • Manage uncertainty as a team	• Maintain self-health • Effective coping with stress • Maintain social support system • Use available resources effectively	Stress and coping fidelity of protocol addressed

Note. QoL = quality of life; RNs = registered nurses.

TABLE 3. Example of Coding for Sample Characteristics

Study	Caregivers With Baseline Data	Enrollment Rate	Patient's Cancer Type and Stage	Caregiver Gender	Caregiver Race	Caregiver Age	Caregiver Attrition
Northouse et al. (2007)	263	69%	Prostate 65% localized stage 21% advanced stage 14% rising PSA stage	99% female	84% White 14% African American 2% Others	59 years	17%

Note. PSA = prostate-specific antigen.

studies as well as from experimental and quasi-experimental studies. The objective of a systematic review is to present the extracted data and summarize them in a meaningful way. For example, Stenberg, Ruland, and Miaskowski (2010) conducted a systematic literature review on the effects of caring for a patient with cancer (see Table 4). They organized findings from their review into a list of five problems and responsibilities reported by family caregivers. They also list the citations that provide evidence for the problem or responsibility identified.

A meta-analysis is a quantitative approach in which data from multiple studies are combined for statistical pooling and generate pooled estimates of intervention effects or estimates of association strength. Meta-analysis presents findings from each primary study in the form of *effect sizes*. An effect size represents the magnitude and the direction of a statistical relationship; it constitutes a variable that is sensitive to differences in strength among studies and represents a common metric that allows researchers to synthesize findings from heterogeneous studies (Lipsey & Wilson, 2001). The effect size statistics that are widely used in meta-analysis are the standardized mean differences and odds ratios. The product-moment correlation coefficient (*r*) constitutes the effect size of the association between two variables. Statistical reports that cannot be used in a meta-analysis are generated by multivariate analysis (e.g., multiple regression, factor analysis, structural equation modeling, etc.; Borenstein et al., 2009; Lipsey & Wilson, 2001). Effect sizes are characterized as *small* (.20), *medium* (.50), or *large* (.80; Cohen, 1988).

There are benefits to presenting study findings in the form of effect sizes. Meta-analysis estimates the effect size for individual studies and pools these estimates across studies. Thus, a meta-analysis produces an estimate of effect sizes with more statistical power than is possible with one individual study. Meta-analysis

TABLE 4. Example of Outcomes of a Systematic Review: Number and References of Types of Problems and Responsibilities Experienced by Family Caregivers of Patients With Cancer

Problems and Responsibilities	Number	and References
Physical, social, or emotional problems and responsibilities in combination	97	3, 8, 9, 25, 42–45, 47, 49, 50, 52, 53, 55, 56, 58–60, 68, 74, 75, 81, 82, 85, 87, 89–157
Social problems and need for information	29	10, 41, 51, 61–63, 66, 67, 69, 86, 88, 119, 158–174
Responsibilities and impact on daily life	20	64, 65, 76–80, 175–187
Emotional problems	11	71, 72, 188–196
Physical health problems and quality of life	7	46, 48, 197–207

Note. Reproduced from "Review of the Literature on the Effects of Caring for a Patient With Cancer," by U. Stenberg, C. M. Ruland, and C. Miaskowski, 2010, *Psycho-Oncology, 19*(10), p. 1015 with permission from John Wiley & Sons, Inc.

can overcome some of the limitations of single studies that have small sample sizes and insufficient statistical power. For example, a meta-analysis examined the effect of interventions targeting caregivers of patients with cancer (Northouse et al., 2010). Among 16 studies, only one had an adequate sample size to detect a significant change in anxiety levels among caregivers randomized to the intervention arm. However, the statistical pooling of data from multiple studies produced a combined sample of $N = 1{,}119$ caregivers. Using this larger sample, Northouse and colleagues found evidence that psychoeducational interventions can reduce anxiety among caregivers of patients with cancer.

Another benefit of a meta-analysis is that *moderator analyses* can be conducted. Moderator analyses identify whether differences in treatment effects are related to characteristics of the sample or the intervention. For example, a meta-analysis by Conn, Hafdahl, Cooper, Ruppar, and colleagues (2009) examined the effects of interventions designed to increase medication adherence among older adults with chronic conditions. Moderation analyses revealed that interventions were more effective for women, for patients taking between three and five medications per day, and when adherence was measured with pill count.

Despite these advantages, meta-analyses also have disadvantages compared with systematic reviews. Meta-analyses apply only to empirical studies that use quantitative measurements and report descriptive and specific inferential statistics. Thus, other types of studies that might produce evidence relevant to practice are excluded. Meta-analyses also require a high level of statistical expertise to calculate effect sizes and pooling data statistically (Borenstein et al., 2009; Lipsey & Wilson, 2001).

Methods for Reporting Results and Assessing for Bias

When findings across studies are homogeneous and consistent (e.g., primary studies reported efficacious interventions with consistent effects), the discussion focuses on the presentation of the summary effect and its meaning for future research and clinical practice. When findings are heterogeneous (e.g., primary reports had conflicting findings and effect sizes vary from small to large), the summary effect becomes less significant and the discussion focuses on the variation among primary studies and possible reasons for this variation (Borenstein et al., 2009).

In systematic reviews, study heterogeneity is best presented by tabulation, which allows readers to look at the methodological rigor of the evidence and the differences between studies. An example of data tabulation is the systematic review and meta-analysis by Katapodi and colleagues (2004; see Figure 2). Data tabulation made evident that using different recruitment methods and various scales to measure perceived breast cancer risk produced distinct differences in effect sizes of perceived breast cancer risk.

In meta-analyses, study heterogeneity is best presented graphically with *forest plots*. A forest plot consists of a perpendicular line that represents the "line of no effect" and numerous horizontal lines. Each horizontal line represents a primary study included in the meta-analysis; each study and the summary effect are depicted as a point estimate bounded by its confidence intervals. A horizontal line to the right of the "no effect line"

	Type of Measurement			
	Numerical		Verbal	
Recruitment	Findings		Findings	
Affected family member, genetic counseling clinic	Daly et al. Dolan et al. Meiser et al. Metcalfe & Narod	Overestimation Overestimation Overestimation Overestimation	Evans et al. Foster et al.	Overestimation Overestimation
Total *N*	1,914		520	
Effect size (95% CI)	+1.26 (+1.19 to +1.33)		+1.14 (+1.00 to +1.27)	
Community setting	Clarke et al. Lipkus et al.	Underestimation Overestimation	Absetz et al. Aiken et al. Facione et al. Lipkus et al. McDonald et al.	Underestimation Underestimation Underestimation Underestimation Underestimation
Total *N*	745		2,963	
Effect size (95% CI)	+2.04 (+1.92 to +2.17)		+0.76 (+0.71 to +0.81)	

Figure 2. Example of table presenting tabulated data.
Note. CI = confidence interval. Reprinted from *Preventive Medicine, 38*(4), "Predictors of Perceived Breast Cancer Risk and the Relation Between Perceived Risk and Breast Cancer Screening: A Meta-Analytic Review," by M. C. Katapodi, K. A. Lee, N. C. Facione, and M. J. Dodd, p. 394, 2004, with permission from Elsevier.

indicates an outcome that favors the intervention, whereas the opposite is true for horizontal lines to the left of the "no effect line." The relative length of each horizontal line represents the 95% confidence intervals of the study's findings; underpowered studies have longer horizontal lines. A forest plot shows if the overall effect is based on many studies or a few, whether studies had adequate statistical power, whether there is significant variation among effects of primary studies, and whether there are outliers that require attention. Thus, a forest plot provides useful information that is easily understood even by people who do not have expertise in meta-analysis and allows readers to evaluate the clinical effect of the treatment (Borenstein et al., 2009).

For example, the meta-analysis by Northouse and colleagues (2010; see Figure 3) included 16 studies that addressed caregivers' anxiety. A forest plot shows that the 16 primary studies reported heterogeneous findings and that most lacked statistical power to detect significant differences in anxiety among caregivers randomized to the intervention arm. The same forest plot shows that although fewer studies

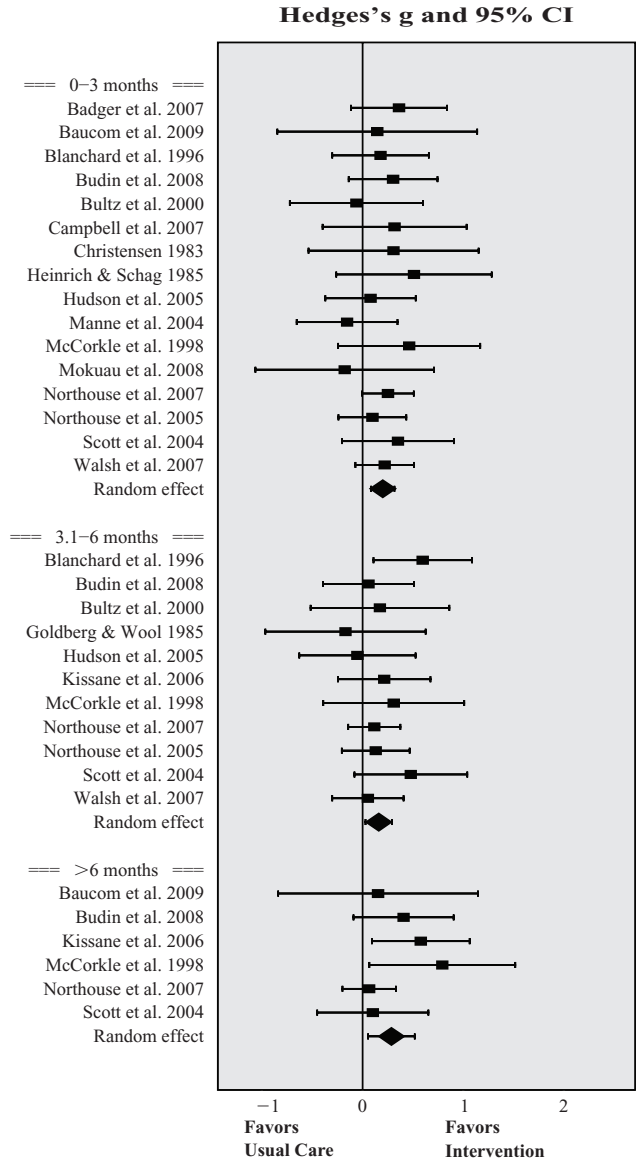

Figure 3. Forrest plot showing effect sizes for anxiety.
Note. CI = confidence interval. Reproduced from "Interventions With Family Caregivers of Cancer Patients: A Meta-Analysis of Randomized Trials," by L. L. Northouse, M. C. Katapodi, L. Song, L. Zhang, and D. W. Mood, 2010, *CA: A Cancer Journal for Clinicians, 60*(5), p. 332 with permission from John Wiley & Sons, Inc.

assessed caregivers' anxiety longer than 3 months postintervention, the overall effect remained small but significant.

When reporting results of a systematic review or meta-analysis, it is essential to assess for bias across studies. The most common source of bias is *publication bias*, resulting to an upward bias or overly positive evaluation of treatment effectiveness

(Lipsey & Wilson, 2001). Publication bias can occur for many reasons. Journals are less likely to publish studies with negative or nonsignificant results, authors are less likely to report negative or inconclusive outcomes from a multi-outcome trial, and studies with small samples are more likely to detect large effect sizes caused by sampling error. Moreover, investigators conducting systematic reviews and meta-analyses often use data only from published studies; this is done because of convenience and with the justification that published material is peer reviewed and of higher quality. However, excluding unpublished studies is likely to introduce publication bias.

Heterogeneity across studies and publication bias can be statistically evaluated with the *Q test*, which is a form of chi-square test, and with *Egger's test*, which is a form of regression (Borenstein et al., 2009). A visual way to assess for publication bias is to create a *scatter plot* of the effect from each study against a measure of its precision. In the absence of bias, the plot should resemble a "funnel shape," which indicates that the desired heterogeneity among studies is adequately represented (e.g., smaller studies and studies with inconclusive or negative results). An asymmetric funnel plot suggests (a) the possibility of publication bias, as smaller studies reporting negative results will be missing; (b) a systematic difference between smaller and larger studies, as smaller, less precise studies are more subject to random variation than larger studies; and (c) use of an inappropriate effect measure (Borenstein et al. 2009). For example, a meta-analysis by Lin and colleagues (2004) examined randomized trials comparing the efficacy of topical NSAIDs with placebo in the management of osteoarthritis pain (see Figure 4). The funnel plot was

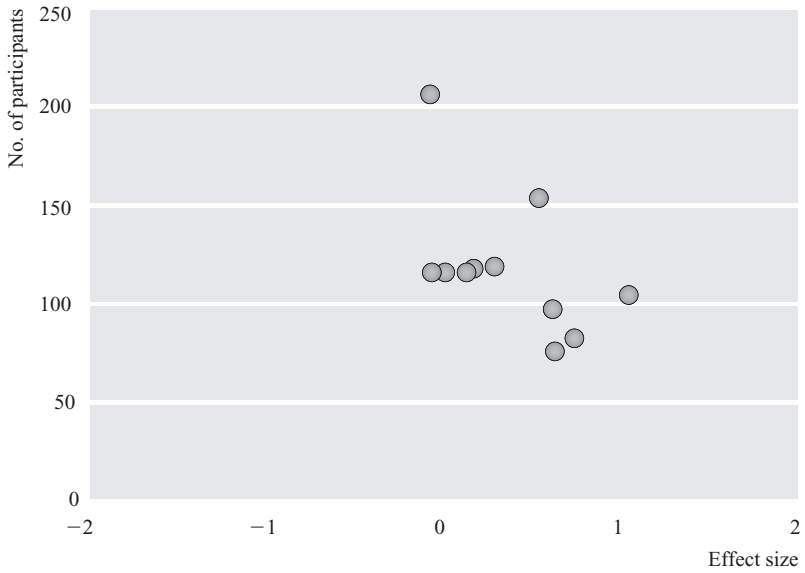

Figure 4. Example of funnel plot: Publication bias. Reproduced from "Efficacy of Topical Non-Steroidal Anti-Inflammatory Drugs in the Treatment of Osteoarthritis: Meta-Analysis of Randomised Controlled Trials," by J. Lin, W. Zhang, A. Jones, and M. Doherty, 2004, *British Medical Journal*, 329(7461), p. 324 with permission from BMJ Publishing Group Ltd.

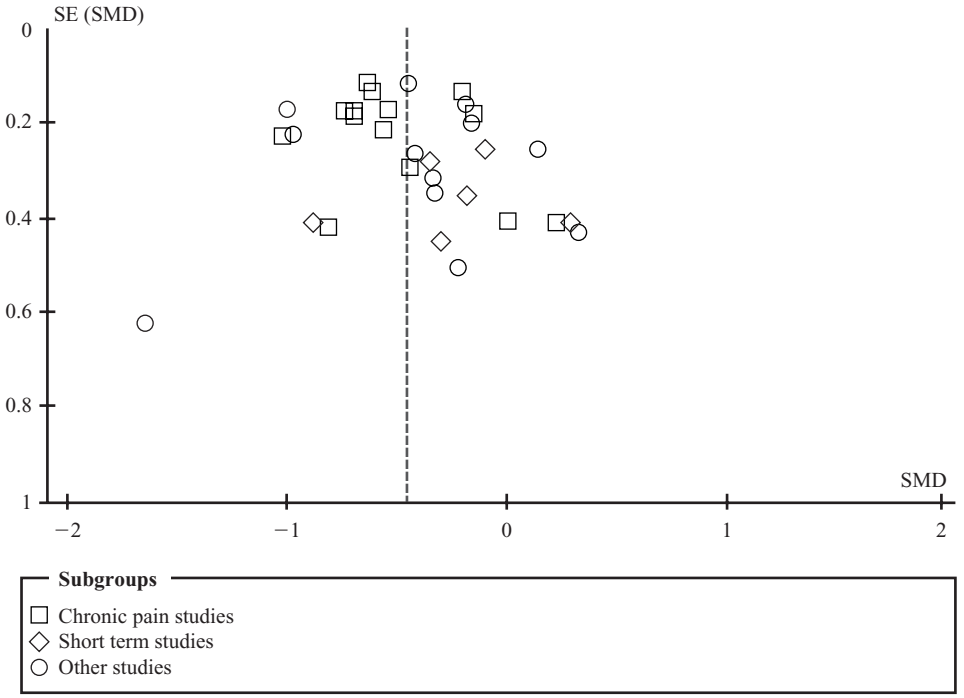

Figure 5. Example of funnel plot: Absence of publication bias. From "How Large Are the Nonspecific Effects of Acupuncture? A Meta-Analysis of Randomized Controlled Trials," by K. Linde, K. Niemann, A. Schneider, and K. Meissner, 2010, *BMC Medicine, 8*, p. 75. Reprinted with permission from BioMed Central.

asymmetrical, indicating that studies with negative results were not included, and that the effect of topical NSAIDs may be overestimated. In contrast, a meta-analysis by Linde, Niemann, Schneider, and Meissner (2010) examined randomized trials of the efficacy of acupuncture over placebo for different illnesses and conditions (see Figure 5). The symmetrical funnel plot indicates absence of publication bias, meaning that smaller studies and studies with negative results were included, and that the effects of acupuncture might be larger than detected.

Interpretation and Clinical Applicability of Findings

In order for systematic reviews and meta-analyses to fully contribute to advancement of evidence-based practice, investigators need to assess the applicability and generalizability of findings. The possibility of harm related to the intervention refers to weighing the likelihood of beneficial effects against the possibility of adverse effects. Interpretation of findings sometimes includes an analysis of the treatment effect per number of patients and the number needed to treat. The number needed to treat examines the overall number of patients that need to be treated with the intervention to see one positive result. Consideration needs to be given to the spectrum of clinical circumstances to which the evidence is likely

applicable. Clinicians need to be cautious not to assume that their own clinical circumstances are similar to those reflected in the included studies (Sidani & Braden, 1998).

CONCLUSIONS AND FUTURE RESEARCH

The call for CER highlights the responsibility to use the best available evidence to inform health care practice. Systematic reviews and meta-analyses can provide evidence for safe and effective care and also provide evidence-based recommendations for practice. In this article, we provided examples of how systematic reviews and meta-analyses (a) can be used to inform clinical decision making for efficacious forms of treatment (e.g., topical NSAIDs for treatment of osteoarthritis and acupuncture for different conditions), (b) can identify subgroups of patients more likely to benefit from a specific intervention (e.g., women are more likely to benefit from nicotine replacement therapy), and (c) intervention characteristics that are more likely to produce results (e.g., interventions that use pill count are more efficacious in promoting medication adherence). These are only a few examples for the use of systematic reviews and meta-analyses in CER.

Decisions about implementation of interventions also need to be informed by an economic evaluation of their performance. This information is useful both to clinicians and to policy makers. For example, Goode and colleagues (1991) evaluated the effects of heparin flush versus saline flush on maintaining patency, preventing phlebitis, and increasing duration of peripheral intravenous lines. Saline flush eliminated problems associated with anticoagulation effects and provided significant cost savings of health care dollars. Thus, future CER could examine how findings of systematic reviews or meta-analyses interface with economic decisions to address questions related to intervention cost-effectiveness (i.e., at which threshold value is the intervention most beneficial and cost-effective). Approaches that integrate meta-analysis and economic decision modeling are complicated and require collaboration from multiple experts, yet, provide a large scope for improvement of CER and a comprehensive view for informed health care decision making (Sutton, Cooper, Goodacre, & Stevenson, 2008).

In summary, systematic reviews and meta-analyses are valuable methods for examining and synthesizing data from a number of empirically based studies and for contributing to CER. There are critical steps in the process of conducting these two important methods for data synthesis. It is hoped that novel investigators and those wishing to conduct CER can appreciate what constitutes a systematic review or meta-analysis and adopt a thorough and objective approach to synthesizing available evidence and informing health care decision making.

REFERENCES

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: John Wiley & Sons.

Cepeda-Benito, A., Reynoso, J. T., & Erath, S. (2004). Meta-analysis of the efficacy of nicotine replacement therapy for smoking cessation: Differences between men and women. *Journal of Consulting and Clinical Psychology*, *72*(4), 712–722. http://dx.doi.org/10.1037/0022-006X.72.4.712

Clark, A. M., Haykowsky, M., Kryworuchko, J., MacClure, T., Scott, J., DesMeules, M., et al. (2010). A meta-analysis of randomized control trials of home-based secondary prevention programs for coronary artery disease. *European Journal of Cardiovascular Prevention and Rehabilitation*, *17*(3), 261–270.

Clark, C. E., Smith, L. F., Taylor, R. S., & Campbell, J. L. (2010). Nurse led interventions to improve control of blood pressure in people with hypertension: Systematic review and meta-analysis. *British Medical Journal*, *341*, c3995. http://dx.doi.org/10.1136/bmj.c3995

Cohen, J. (1988). The effect size index: d. In *Statistical power analysis for the behavioral sciences* (2nd ed., pp. 20–26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conn, V. S., Hafdahl, A. R., & Brown, L. M. (2009). Meta-analysis of quality-of-life outcomes from physical activity interventions. *Nursing Research*, *58*(3), 175–183.

Conn, V. S., Hafdahl, A. R., Cooper, P. S., Brown, L. M., & Lusk, S. L. (2009). Meta-analysis of workplace physical activity interventions. *American Journal of Preventive Medicine*, *37*(4), 330–339.

Conn, V. S., Hafdahl, A. R., Cooper, P. S., Ruppar, T. M., Mehr, D. R., & Russell, C. L. (2009). Interventions to improve medication adherence among older adults: Meta-analysis of adherence outcomes among randomized controlled trials. *The Gerontologist*, *49*(4), 447–462.

Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Identifying relevant studies for systematic reviews. *British Medical Journal*, *309*(6964), 1286–1291.

Goode, C. J., Titler, M., Rakel, B., Ones, D. S., Kleiber, C., Small, S., et al. (1991). A meta-analysis of effects of heparin flush and saline flush: Quality and cost implications. *Nursing Research*, *40*(6), 324–330.

Institute of Medicine. (2009). *Initial national priorities for comparative effectiveness research*. Washington, DC: The National Academies Press.

Katapodi, M. C., Lee, K. A., Facione, N. C., & Dodd, M. J. (2004). Predictors of perceived breast cancer risk and the relation between perceived risk and breast cancer screening: A meta-analytic review. *Preventive Medicine*, *38*(4), 388–402. http://dx.doi.org/10.1016/j.ypmed.2003.11.012

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. http://dx.doi.org/10.1371/journal.pmed.1000100

Lin, J., Zhang, W., Jones, A., & Doherty, M. (2004). Efficacy of topical non-steroidal anti-inflammatory drugs in the treatment of osteoarthritis: Meta-analysis of randomised controlled trials. *British Medical Journal*, *329*(7461), 324. http://dx.doi.org/10.1136/bmj.38159.639028.7C

Linde, K., Niemann, K., Schneider, A., & Meissner, K. (2010). How large are the nonspecific effects of acupuncture? A meta-analysis of randomized controlled trials. *BMC Medicine*, *8*, 75.

Lipsey, M. W., & Wilson, D. B. (Eds.). (2001). *Practical meta-analysis* (Applied social research methods series, Vol. 49). Thousand Oaks, CA: Sage.

McCaul, K. D., Branstetter, A. D., Schroeder, D. M., & Glasgow, R. E. (1996). What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. *Health Psychology*, *15*(6), 423–429.

McManus, R. J., Wilson, S., Delaney, B. C., Fitzmaurice, D. A., Hyde, C. J., Tobias, R. S., et al. (1998). Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *British Medical Journal*, *317*(7172), 1562–1563.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097

Northouse, L. L., Katapodi, M. C., Song, L., Zhang, L., & Mood, D. W. (2010). Interventions with family caregivers of cancer patients: A meta-analysis of randomized trials. *CA: A Cancer Journal for Clinicians*, *60*(5), 317–339.

Papadakis, S., McDonald, P. A., Mullen, K. A., Reid, R., Skulsky, K., & Pipe, A. (2010). Strategies to increase the delivery of smoking cessation treatments in primary care settings: A systematic review & meta-analysis. *Preventive Medicine*, *51*(3–4), 199–213. http://dx.doi.org/10.1016/j.ypmed.2010.06.007

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *The Journal of the American Medical Association*, *273*(5), 408–412.

Sellers, D. E., Crawford, S. L., Bullock, K., & McKinlay, J. B. (1997). Understanding the variability in the effectiveness of community heart health programs: A meta-analysis. *Social Science & Medicine*, *44*(9), 1325–1339.

Sidani, S., & Braden, C. J. (1998). *Evaluating nursing interventions: A theory-driven approach*. Thousand Oaks, CA: Sage.

Sox, H. C., & Greenfield, S. (2009). Comparative effectiveness research: A report from the Institute of Medicine. *Annals of Internal Medicine*, *151*(3), 203–205.

Sox, H. C., Helfand, M., Grimshaw, J., Dickersin, K., The PLoS Medicine Editors, Tovey, D., et al. (2010). Comparative effectiveness research: Challenges for medical journals. *PLoS Medicine*, *7*(4), e1000269. http://dx.doi.org/10.1371/journal.pmed.1000269

Stenberg, U., Ruland, C. M., & Miaskowski, C. (2010). Review of the literature on the effects of caring for a patient with cancer. *Psycho-Oncology*, *19*(10), 1013–1025. http://dx.doi.org/10.1002/pon.1670

Sutton, A. J., Cooper, N. J., Goodacre, S., & Stevenson, M. (2008). Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. *Medical Decision Making*, *28*(5), 650–667. http://dx.doi.org/10.1177/0272989X08324036

Taylor-Piliae, R. E., & Froelicher, E. S. (2004). The effectiveness of Tai Chi exercise in improving aerobic capacity: A meta-analysis. *The Journal of Cardiovascular Nursing*, *19*(1), 48–57.

Correspondence regarding this article should be directed to Maria C. Katapodi, PhD, MSc, BSN, University of Michigan School of Nursing, 400 Ingalls Building, Room 2158, Ann Arbor, MI 48109-5482. E-mail: mkatapo@umich.edu