Development of Clinical Practice Guidelines

Steven D. Hollon,[1,10] Patricia A. Areán,[2,10] Michelle G. Craske,[3,10] Kermit A. Crawford,[4,10] Daniel R. Kivlahan,[5,10] Jeffrey J. Magnavita,[6,10] Thomas H. Ollendick,[7,10] Thomas L. Sexton,[8,10] Bonnie Spring,[9,10] Lynn F. Bufka,[11] Daniel I. Galper,[11] and Howard Kurtzman[11]

[1]Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240; email: steven.d.hollon@vanderbilt.edu

[2]Department of Psychiatry, School of Medicine, University of California, San Francisco, California 94143

[3]Department of Psychology, University of California, Los Angeles, California 90095

[4]Center for Multicultural Mental Health, Boston University Medical Campus, Boston, Massachusetts 02118

[5]Department of Psychiatry and Behavioral Sciences, University of Washington and Mental Health Services, Veterans Health Administration, Seattle, Washington 98108

[6]Glastonbury Psychological Associates, Glastonbury, Connecticut 06033

[7]Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

[8]Department of Counseling and Educational Psychology, Indiana University, Bloomington, Indiana 47405

[9]Department of Preventive Medicine, Northwestern University, Chicago, Illinois 60611

[10]Advisory Steering Committee, American Psychological Association, Washington, DC 20002

[11]American Psychological Association, Washington, DC 20002

Annu. Rev. Clin. Psychol. 2014. 10:213–41

The *Annual Review of Clinical Psychology* is online at clinpsy.annualreviews.org

This article's doi: 10.1146/annurev-clinpsy-050212-185529

Keywords

clinical practice guidelines, guideline development panels, systematic reviews, conflict of interest, analytic framework (scoping), PICOTS questions, transparency

Abstract

Clinical practice guidelines (CPGs) are intended to improve mental, behavioral, and physical health by promoting clinical practices that are based on the best available evidence. The American Psychological Association (APA) is committed to generating patient-focused CPGs that are scientifically sound, clinically useful, and informative for psychologists, other health professionals, training programs, policy makers, and the public. The Institute of Medicine (IOM) 2011 standards for generating CPGs represent current best practices in the field. These standards involve multidisciplinary guideline development panels charged with generating recommendations based on comprehensive systematic reviews of the evidence. The IOM standards will guide the APA as it generates CPGs that can be used to inform the general public and the practice community regarding the benefits and harms of various treatment options. CPG recommendations are advisory rather than compulsory. When used appropriately, high-quality guidelines can facilitate shared decision making and identify gaps in knowledge.

Contents

INTRODUCTION

Clinical Practice Guidelines Defined

Evidence-based clinical practice guidelines (CPGs) represent a systematic approach to translating the best available research evidence into clear statements regarding treatments for people with various health conditions. Because health care costs have exploded, our health care system has dual responsibilities: to ensure that individuals are treated according to best practices and to reduce unnecessary expenditures. Quality CPGs based on comprehensive reviews of the research evidence can help patients, practitioners, policy makers, and administrators make better decisions about how to proceed with care. Some have viewed CPGs with skepticism because of their potential for misuse, as when care is restricted inappropriately, but high-quality guidelines, used appropriately, can facilitate shared decision making and identify gaps in knowledge.

Evidence-based CPGs are important tools for making treatment decisions. As shown in **Figure 1**, evidence-based decision making filters (*a*) the best available research evidence through the prism of (*b*) practitioner expertise to arrive at the best decisions consistent with (*c*) patient preferences and values (Spring & Hitchcock 2009). This three-circle model was first proposed in the context of evidence-based medicine to facilitate "the conscientious and judicious use of current best evidence from clinical care research in the management of individual patients" (Sackett et al. 1996, p. 71) and was subsequently adopted with modifications as policy by both the Institute of Medicine (IOM) (2001) and the American Psychological Association (APA) (APA Pres. Task Force Evid. Based Pract. 2006). The model depicted in **Figure 1** reflects the importance of grounding clinical practice in high-quality science (with different research designs suited to different questions) but also reflects the important role played by the clinician in deciding how to adapt that information to the specific needs and preferences of the individual patient. CPGs

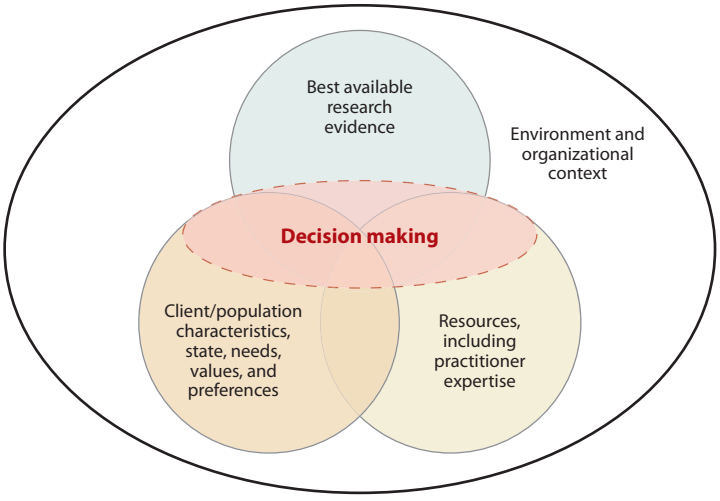


Figure 1

Tripartite model for clinical decision making. Reprinted with permission from Spring & Hitchcock (2009).

facilitate this process by providing actionable recommendations based on evaluating the pros and cons of different treatment options. Any set of guidelines must come to terms with the interplay between the intervention, the setting in which it is delivered, and the patient who receives it. CPGs are intended to assist the clinician in dealing with the inherent complexity of decision making in clinical situations.

It is important to note what guidelines are not. CPGs are not intended to establish standards of care that have legal ramifications. Requirements to report child abuse or warn of imminent harm are standards of care that must be followed under penalty of law. CPGs are only advisory. CPGs are not treatment manuals or cookbooks that tell a clinician how to implement the recommended treatments. They may suggest where to go to learn more about a recommended treatment but will not provide the necessary instructions to implement that intervention. CPGs are not a substitute for good clinical judgment; rather, they are intended to be an aid to the clinician and patient. Both the IOM (2001) and the APA (APA Pres. Task Force Evid. Based Pract. 2006) explicitly recognize the central role of clinical expertise in the process of selecting and implementing treatments. CPGs are not the sole determinants of treatment plans; rather, they inform the general public and the practice community about the range of options available and their respective strengths. CPGs do not set reimbursement policies; rather, they provide a bulwark against policies that seek to subjugate good clinical practice to economic considerations (APA 2002). Finally, CPGs do not constrain practice; rather, they seek to inform and enhance it. The treating clinician is always free not to follow specific recommendations when there are good clinical reasons not to do so.

Examples of high-quality CPGs include preventive care guidelines produced by the US Preventive Services Task Force, tobacco treatment guidelines prepared by the US Public Health Services, and 1998 obesity treatment guidelines developed by the National Heart, Lung, and Blood Institutes. The Department of Veterans Affairs (VA) and the Department of Defense (DoD) generate joint CPGs relevant to health care providers in the Veterans Health Administration and Military Health, but those guidelines are not always relevant to civilian populations. Agencies in other countries, such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom, generate high-quality CPGs based on systematic literature reviews; the

recommendations generated there are not always relevant to the US health care system. Many of the CPGs developed in this country are produced by professional societies or proprietary organizations. The former risks bias in favor of the services offered by that particular profession; the latter risks bias when the profit motive takes precedence over the public interest. **Table 1** provides a list of recent mental and behavioral health guidelines.

Table 1 Recent mental and behavioral health clinical practice guidelines

	USPSTF	VA/DoD	NICE	SIGN	AACAP	ApA	Cin Child	ICSI
Anxiety Disorders								
ASD/PTSD		2010	2005			2004		
OCD			2005			2007		
Panic disorder			2011			2009		
Childhood/adolescence disorders								
ADHD			2008	2009	2007		2009	2012
Anxiety disorders					2007			
Autism spectrum			2012	2007				
Bipolar disorder					2007			
Depression (MDD)	2009		2005		2007		2010	2010
Eating disorders								
ODD					2012			
Eating disorders/obesity								
Eating disorders			2004			2006		
Obesity	2012	2006	2006	2010				2011
Mood disorders								
Bipolar disorder		2010	2006	2005		2002		
Depression (MDD)		2009	2009	2010		2010		2012
Computerized			2006					
Organic brain damage								
Dementia			2006	2006		2007		
TBI/stroke		2009	2007	2010			2012	
Personality disorders								
Antisocial			2009					
Borderline			2009			2001		
Pain (chronic)		2010						
Schizophrenia			2009			2004		
Sexual dysfunction								
Sleep disorders								
Substance abuse		2009	2011	2003		2006		

Abbreviations: AACAP, American Academy of Child and Adolescent Psychiatry; ADHD, attention deficit hyperactivity disorder; ApA, American Psychiatric Association; ASD, acute stress disorder; Cin Child, Cincinnati Children's Hospital Medical Center; ICSI, Institute for Clinical Systems Improvement; MDD, major depressive disorder; NICE, National Institute for Health and Care Excellence (UK); OCD, obsessive-compulsive disorder; ODD, oppositional defiant disorder; PTSD, posttraumatic stress disorder; SIGN, Scottish Intercollegiate Guidelines Network (UK); TBI, traumatic brain injury; USPSTF, United States Preventive Services Task Force; VA/DoD, Veterans Affairs/Department of Defense.

Recognizing a need for unbiased and evidence-based CPGs for use in the United States, the APA has decided to generate clinical practice guidelines to serve the larger public interest. For this purpose, it formed an advisory steering committee (ASC) (the authors of this article) to formulate recommendations regarding how to go about the development process. We are still in the process, and not all of our decisions are final. We have immersed ourselves in literature regarding guideline development and have talked extensively with other organizations that have undertaken similar endeavors. We have drawn heavily on the expertise of members involved in the generation of the VA/DoD guidelines and are particularly indebted to the individuals involved in generating the NICE guidelines for sharing both their time and their expertise.

We have been strongly influenced by standards recently published by the IOM for generating trustworthy CPGs (2011a) based on independent and comprehensive systematic reviews (SRs) of the literature (2011b). These standards provide expert consensus on a compilation of the current best practices in guideline generation in the world today, and we have strongly encouraged the APA to follow those recommendations. We have tried to define a process that serves the public interest rather than the vested interests of psychology as a profession. Our guiding principle has been that if what we do is good for the general public, it will be good for psychology as a discipline. Our goal is to develop CPGs that the general public and clinical practitioners can use to inform their treatment decisions and that mental health administrators can use to promote access to an appropriate range of treatment options. **Table 2** describes our mission as we see it, the vision that we followed, and the operational principles we adopted to guide our efforts.

We recognize the complexity of the treatment process. There may be factors specific to a given intervention, nonspecific (common) factors shared across interventions, and therapist factors that together compose the psychosocial therapies. All of the factors can contribute to the clinical change process, and they can do so in a myriad of interactive ways. Moreover, different patients may respond to different aspects. CPGs can identify what works and highlight aspects of inquiry into these diverse components that need further exploration.

Policy Shift for the APA

The APA's entry into the arena of CPG development represents a significant evolution in policy development. Until recently, the APA produced only general guidelines for professional conduct and certain aspects of psychological practice (which it called practice guidelines). In February 2010, however, it endorsed a process for developing treatment guidelines that provide recommendations about specific clinical interventions for specific problem areas. To align its terminology with the broader health care arena, in August 2012, the APA revised its nomenclature such that guidelines specific to treatment are now called CPGs, whereas other guidelines are professional practice guidelines.

The change in terminology has brought the organization's vocabulary in line with the rest of the health care field and reflects a significant evolution of policy at the APA with respect to CPGs. In the 1990s, APA's Board of Professional Affairs, Board of Scientific Affairs, and Committee for the Advancement of Professional Practice determined that the organization would not create CPGs but would instead develop a policy that could be used to evaluate guidelines developed by other organizations. This decision resulted in the *Template for Developing Guidelines: Interventions for Mental Disorders and Psychosocial Aspects of Physical Disorders* (APA 1995), approved by the APA Council of Representatives in February 1995. The document was subsequently revised, and in 2000 the APA Council of Representatives approved the *Criteria for Evaluating Treatment Guidelines* as policy (APA 2002). These latter criteria provide a systematic, APA-approved policy to identify

Table 2 Mission, vision, and operational principles of the American Psychological Association's Advisory Steering Committee

Mission: Improve mental, behavioral, and physical health by promoting clinical practices based on the best available evidence. Identify interventions that are effective and can be implemented in the community. Develop treatment guidelines that are scientifically sound, clinically useful, and informative for psychologists, other health professionals, training programs, policy makers, and the public.
Vision
Improve mental, behavioral and physical health.
Improve patient experiences of care.
Improve the effectiveness, quality, and value of health care services.
Inform shared decision making between patients and health professionals.
Improve practice by health professionals.
Enhance training of psychologists and other health professionals.
Enhance competency of psychologists and other health professionals.
Educate professionals, consumers, and policy makers about effective interventions.
Identify gaps in the evidence base to be addressed by future research.
Operational principles
Focus on treatment efficacy and clinical utility for specific problems or disorders.
Utilize transparent rationale and procedures for guideline development.
Solicit input from professionals, consumers, educators, payers, and policy makers.
Focus on prevention, treatment, or management.
Address common factors and relational issues associated with effective interventions.
Consider the full range of research evidence.
Evaluate the quality of the evidence and identify areas where it is lacking.
Identify potential harms and benefits.
Consider outcomes across multiple domains.
Consider setting, sociodemographics, multicultural issues, and patient preferences.
Acknowledge the clinician's responsibility for individualized clinical care decisions.
Provide recommendations that are advisory rather than compulsory.
Update guidelines periodically to reflect developments in research and practice.

and delimit the promulgation of poor, unreasonable, unjustified, or biased treatment guidelines in the marketplace (see Stricker et al. 1999).

In 2004, APA president-elect, Ron Levant, PhD, appointed the Presidential Task Force on Evidence-Based Practice and directed its members to develop a policy statement for the association. The resulting statement defined evidence-based practice in psychology as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Pres. Task Force Evid. Based Pract. 2006, p. 273). The statement was adopted as APA policy in August 2005, and an accompanying report elaborating on these principles was subsequently published (APA Pres. Task Force Evid. Based Pract. 2006). These documents underscored that psychological practice is built on a foundation of research but also generated controversy: Some think the documents undervalue research and others that they do not sufficiently value clinical expertise.

Shortly after this policy statement and report were developed, governance groups and key leaders in psychology began to question the wisdom of the APA's decision to refrain from

developing CPGs. A prominent concern was that noninvolvement in CPG development would leave the evaluation of evidence to other developers with conflicting interests or lack of familiarity with the full range of relevant psychological science. Concern that existing guidelines overemphasize medications to the relative neglect of often more preferred and effective psychosocial interventions gradually prompted the realization that guideline development by the APA was necessary. CPGs afford an opportunity to use psychological science to support clinical practice. Historically, psychologists are well trained in research and hence have the skills to understand and oversee the systematic reviews necessary to develop sound guidelines. For all of these reasons, the APA ultimately decided to undertake a pilot project in guideline development.

GENERATING CLINICAL PRACTICE GUIDELINES

Guideline Generation and the Nature of the Evidence

Both IOM (2001) and APA (APA Pres. Task Force Evid. Based Pract. 2006) policies acknowledge that decision making is too complex to rest solely on empirically based prescriptive recommendations and too consequential to rely only on clinical consensus. CPG generation typically involves two steps: (*a*) the conduct of a comprehensive SR of the empirical literature that is presented to a guideline development panel (GDP) and (*b*) evaluation of the findings from the SR by the GDP that considers the quality of evidence and the relative benefits and harms associated with the clinical practices reviewed. The GDP, consisting of practitioners, scientists, methodological experts, and patient representatives, is then asked to generate recommendations that are informed by the empirical literature but take clinical experience into account, as depicted in **Figure 2**. In consequence, any given CPG is grounded in the best available research evidence as filtered through expert clinical judgment and with consideration of patient preferences. The process is like a jury trial in which conflicting evidence is presented to an independent group charged with arriving at a set of final decisions.

Exactly what constitutes evidence for the SR depends on the nature of the question being asked. In most instances our GDPs will be asking which treatments work best (efficacy) and for whom (called moderation by psychologists and effect modification by epidemiologists). To make causal inferences about treatment efficacy, most efforts at evidence-based medicine assume a hierarchy of evidence that stretches from clinical consensus up to rigorous and well-implemented randomized controlled trials (RCTs). Quasi-experimental designs and benchmarking studies (Minami et al. 2007) can be useful for determining whether a treatment's effects generalize to real-world settings (effectiveness). Cohort designs stand at the top of the hierarchy for questions about prognosis, and case studies or qualitative designs may be best suited for exploring a novel phenomenon or context (Vandenbroucke 2008).

RCTs are often referred to as the gold standard for determining whether a treatment works (has a causal impact) because they do a better job of ruling out rival plausible explanations for the observed outcomes (such as selection bias) than other designs. Nonetheless, as guides for clinical decision making, they do have some limitations (Westen et al. 2004). For example, the fact that something works on average does not mean that it will work for any given patient. Most RCTs select samples based on diagnosis, but that may not be the only (or even the most) important patient characteristic to consider. Any of a variety of patient characteristics may play a role in determining the best treatment for a given individual (moderation), and it is the clinician who is charged with weighing the available information in each case to determine (in consultation with the patient) the most promising line of intervention. Tests of moderation can be built into RCTs and often are, but we still know far too little about how to weight the complex factors involved

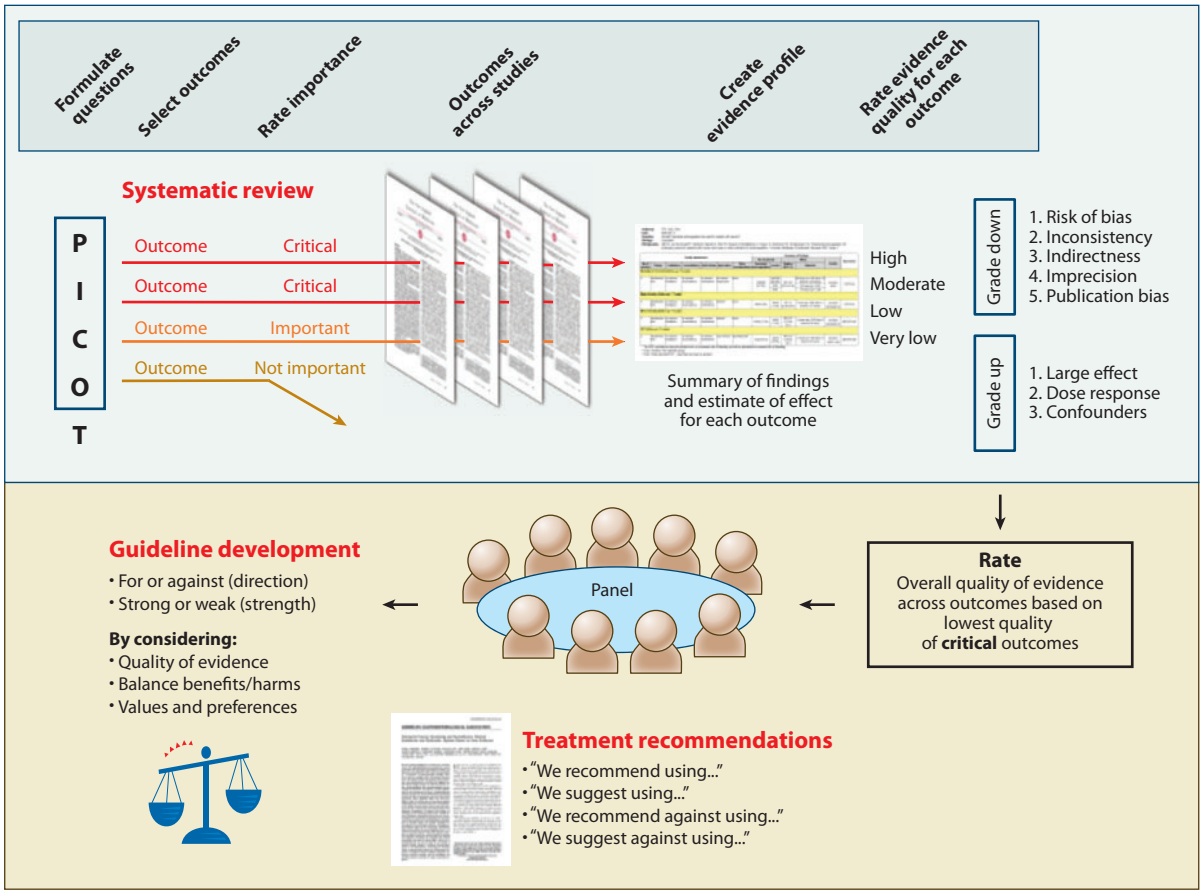


Figure 2

Systematic reviews used to inform guideline development panels. Reprinted with permission from Falck-Ytter & Schünemann (2009).

in determining the best treatment for a given patient. Well-constructed guidelines can highlight what is currently known in that regard.

Moreover, treatments found to be efficacious in RCTs and recommended by guidelines often have only a probabilistic causal relationship to improved clinical outcomes. Probabilistic causality tells us only that the average patient who receives the treatment is likely to respond more favorably than if he or she receives no treatment or a less potent one; it does not tell us why (mediation). To increase the proportion of respondents and especially the magnitude of response requires pushing the frontiers of intervention science beyond probabilistic and toward mechanistic causality (Borgerson 2009). Mechanistic causal understanding represents a higher standard, reflecting an ability to map the influences and steps in the causal chain whereby the treatment drives the clinical outcome. Tests of moderation and mediation can be carried out within the context of RCTs (Kraemer et al. 2002), and tests of mediation can be made even more powerful when the purported mechanism can be manipulated independent of treatment (Maier et al. 2006). CPGs can document who benefits from and the mechanisms underlying any given treatment so as to increase both efficiency (moderation) and intervention power (mediation).

Clinicians have an opportunity to advance causal understanding and personalize treatment by treating clinical encounters as experiments intended to optimize treatment for an individual client. Unless there are strong contraindications, the evidence-based practice process (**http://www.ebbp.org**) usually begins by applying a somewhat generic treatment shown to be effective for the average individual. The therapist then assesses the patient's response and experiments with technical or relational changes to gauge whether these heighten a positive response. By analyzing the patient's response and adjusting treatment accordingly, the clinician iterates back and forth, tailoring a generic treatment to the individual patient. Insights gained from this process of individualizing treatment and analyzing responses can be used to generate hypotheses about causal mechanisms that can then be tested in a more systematic fashion by using single-case experimental, intermittent time series, or clinical trial designs. This process invites clinicians to contribute to the advancement of the next generation of intervention science by playing a role in the context of discovery while simultaneously recognizing the need to incorporate systematic controls to rule out plausible rival causal explanations (the context of verification).

This approach acknowledges the need to improve our understanding of what mediates clinical improvement, which likely includes causally active treatment processes (specific or nonspecific and often relational) (Magnavita 2010, Sexton et al. 2011), as well as processes of change that are mobilized in patients and the treatment's compatibility with the patient (Beutler et al. 2012). It also suggests a way that nomothetic information derived from RCTs can be shaped to fit the individual patient. We endorse the notion that many different kinds of information are needed to continue to improve treatments in order to increase the number of people helped and the degree of improvement our treatments bring about. Our goal is to capitalize on all research that helps enhance clinical understanding and improve client outcomes while recognizing that different designs are best suited to address different kinds of questions.

Guideline Development Panels

Prior CPGs have been quite diverse with respect to the methods followed and the way their recommendations were made. The recently published IOM standards promise to address persistent deficiencies in the way CPGs are developed (Kung et al. 2012). One standard deals with how GDPs charged with generating the treatment recommendations are formulated and function (IOM 2011a). The other deals with the execution of SRs of the literature on which those recommendations are based (IOM 2011b). These standards with respect to GDPs are listed in **Table 3** and discussed below.

Standard 1: Establishing transparency. Transparency means that sufficient information is provided to allow guideline users to understand the way the recommendations were derived. We plan to post the processes followed in guideline development on a public website, along with the means by which GDP members are recruited and the selection process involved. Potential conflicts of interest (COIs) also will be posted so that guideline users can evaluate the perspectives that the panelists brought to bear. The APA will provide funding for the initial set of guidelines, and we hope to share funding for joint guidelines with other professional organizations in the future. The goal is to be as transparent as possible about the developmental process so that others can determine how much they can trust the resultant recommendations.

Standard 2: Management of conflict of interest. There is potential for COIs given any "set of circumstances that creates a risk that professional judgment or actions regarding a primary interest will be unduly influenced by a secondary interest" (IOM 2011a, p. 46). This includes not only opportunities for financial gain or professional advancement but also the intellectual passions of

Table 3 Institute of Medicine standards for guideline generation

Standard	Description
Establishing transparency	Standard 1: The processes by which a CPG is developed and funded should be described explicitly and be publicly accessible.
Management of COI	Standard 2: Potential panel members should disclose all potential COIs prior to selection and within the GDP and be excluded from votes relevant to those conflicts, with divestment recommended.
GDP composition	Standard 3: Composition should be multidisciplinary and balanced with respect to orientation and clinical/research expertise; potential patient and lay representatives should be involved at some point in the guideline development process.
CPG–SR interaction	Standard 4: CPGs are based on SR of literature with specific interaction between the GDP and SR team.
Establishing evidence base and rating recommendation strength	Standard 5: Each recommendation should be accompanied by a clear description of the benefits and harms and a summary of the available evidence (and evidentiary gaps) along with ratings of the quality of the evidence underpinning the recommendation and the strength of the recommendation (with differences of opinion documented).
Articulation of recommendations	Standard 6: Recommendations should be stated in standardized format that describes precisely what the recommended action is and under what circumstances it should be performed, so compliance can be evaluated.
External review	Standard 7: Guidelines should be made available for review prior to publication, with comments kept confidential and written responses made to each.
Updating	Standard 8: Guidelines should be updated periodically as new evidence emerges.

Abbreviations: COI, conflict of interest; CPG, clinical practice guideline; GDP, guideline development panel; SR, systematic review. From Inst. Med. (2011a).

the panelists. Prospective panelists and members of the advisory steering committee will be asked to disclose potential COIs as part of the selection process and to do so again at the initial meeting to ensure that they are aware of one another's conflicts. All those involved in generation of SRs or CPGs will be asked to initially disclose and periodically update information about their COIs. All COIs will be published in the final guidelines.

Potential GDP members with COIs will not necessarily be disqualified from service on a panel nor asked to divest themselves of relevant financial interests. COIs often go hand-in-hand with the expertise necessary to generate evidence-based CPGs. Rather than risk minimizing expertise, we plan to follow the principle of adversarial collaboration, in which competing interests are balanced rather than excluded (Mellers et al. 2001). Members with COIs will be allowed to participate in all discussions but will not be allowed to chair the GDP or to vote on specific recommendations regarding their area of conflict. Requiring that all COIs be disclosed and balancing competing perspectives should allow the weight of the evidence channeled through the group consensus to carry the day.

Standard 3: Guideline development panel composition. As recommended by the IOM, we plan to form multidisciplinary GDPs. Group composition influences the nature of the recommendations, and members of a clinical specialty are likely to promote interventions relevant

to their specialty. Multidisciplinary groups tend to be more conservative in their recommendations than groups drawn from a single discipline, and different specialty groups tend to arrive at contrasting conclusions weighted in the direction of their respective practices. We therefore strongly endorse the principle of multidisciplinary panels and plan to make every effort to include representatives of key specialties involved in the treatment of the target disorder.

The first GDP that we constituted was on the treatment of unipolar depression across the life span. Depression is now twice as likely to be treated with medications as with psychotherapy (Marcus & Olfson 2010). This represents a total reversal in the relative frequencies with which the two approaches were used over the last two decades, with much of the increase in medication usage seen in general practice (Olfson et al. 2002). Given the widespread use of antidepressant medications and the frequency with which they are prescribed in primary care, we included 3 physicians on the 13-member depression GDP, including a general practitioner with expertise in family systems. One of the 2 psychiatrists is an expert in the treatment of geriatric populations, and the other is the chief medical officer of a large managed care organization.

We also wanted the GDP to be diverse with respect to theoretical orientation (at least one panelist each is associated with dynamic, interpersonal, behavioral, cognitive, experiential, and family systems approaches) and professional activity (most see patients and two are engaged in full-time practice or administration). One panelist is a research methodologist with expertise in meta-analysis. Two of the panelists are specialists in child and adolescent depression, and two others are specialists in geriatric depression. We also wanted the panel to be diverse with respect to ethnicity and race (two panelists are of Asian/Pacific Islander ancestry and three of black/African ancestry; the remaining nine panelists are Caucasian) and gender (four are female, eight are male, and one is transgender). Two of the panelists are from other countries (one from Canada and the other from the Netherlands).

The IOM also recommends including patients or lay representatives in the guideline development process, although opinion remains divided as to exactly how to do that. Patients and lay representatives can provide information on outcomes of interest to consumers and act as a safeguard against COIs that may bias the judgment of clinical or research experts, but some can be too attached to whatever treatment they received, and some advocacy groups are largely funded by commercial interests. After some debate, we decided to include two community members on the GDP and to seek input from the general public and patient advocacy groups during an open comment period so that we could integrate their comments on the preliminary draft into the final report.

Standard 4: Interaction between the GDP and the SR team. For our first several CPGs, we intend to contract with independent groups experienced in conducting SRs according to current best practices (IOM 2011b). The IOM describes several different models of interaction between the GDP and the SR team ranging from complete isolation through full interaction. The former minimizes the risk of bias but does so at the cost of efficiency. We plan to follow the recommended limited interaction model, in which members of the SR team initially meet with the GDP to define the scope of the review and then return on an as-needed basis to explain the nature of the findings and to respond to requests for additional information.

Standard 5: Rating strength of recommendations. CPGs essentially rest on an appraisal of the quality of the relevant evidence, a comparison of the benefits and harms of particular clinical recommendations, and value judgments regarding the importance of specific benefits and harms. Several strategies have been developed for making explicit the strength of recommendations. Perhaps the most widely used system is Grading of Recommendations Assessment, Development and Evaluation (GRADE) (Schünemann et al. 2009). Although it has many desirable features, it can be cumbersome to use. The American College of Chest Physicians (ACCP) has developed a

streamlined version of GRADE for appraising evidence quality and strength of recommendations (Guyatt et al. 2006). Strong recommendations indicate that the benefits clearly outweigh the harms (or vice versa), whereas weak recommendations indicate a less clear balance between benefits and harms. APA will likely similarly modify GRADE for its own use by our CPGs.

Standard 6: Articulation of recommendations. Guidelines need to be clear about what actions they are recommending. Recommendations that are vague or nonspecific are likely to be misinterpreted and unlikely to be followed. We plan to ask the GDPs to generate key action statements that recommend specific clinician behavior. The IOM recommends semantic separation of strong from weak recommendations ("we recommend" or "clinicians should" versus "we conditionally recommend" or "clinicians might"). Some recommendations provide little actionable guidance and make it difficult to determine whether they were implemented (e.g., "providers should consider. . ."). We have not yet settled on the language we will use, but we do plan to adopt a consistent format across all our CPGs. In the absence of high-quality empirical evidence, recommendations based on clinical consensus must be clearly identified.

Standard 7: External review. We plan to solicit comments from selected external reviewers and to share drafts of the CPG with the general public during comment periods prior to publication. These reviewers can call attention to errors and omissions and can provide valuable input for improving the recommendations. Relevant reviewers may include research and clinical experts in the area, representatives of federal agencies or professional organizations, representatives of health plans, and representatives of advocacy organizations and the general public. Consistent with IOM standards, we plan to keep authorship of the critiques confidential so that the reviewers will feel free to express their opinions. In the interest of transparency, we also plan to keep a written record of each comment and any change that is made in response.

Standard 8: Updating. CPG recommendations require updating when there are significant changes in (*a*) evidence of benefits and harms, (*b*) outcomes that are considered important, (*c*) available interventions, (*d*) evidence that current practice is optimal, (*e*) the value placed on different outcomes, or (*f*) resources available for health care. The ASC will monitor each guideline in an ongoing fashion and determine when an update is required in accordance with recent guidelines from the Agency for Healthcare Research and Quality (Newberry et al. 2013).

Conducting Systematic Reviews

The IOM standards for conducting SRs represent current best practice in the field, with an emphasis on transparency (IOM 2011b). Each standard includes a set of requirements to guide performance. As shown in **Table 4**, the IOM has developed four sets of standards with respect to stages of conducting SRs: (*a*) initiating a review, (*b*) finding and assessing individual studies, (*c*) synthesizing the body of evidence, and (*d*) reporting the results. We discuss each in turn.

Initiating the systematic review. The IOM calls for assembling a team with the requisite skills to conduct the SR (including expertise in information search and quantitative methods) and for soliciting input from guideline users and stakeholders. It also calls for managing bias or potential COI in the SR team, with disclosure being paramount. Standards are provided for developing the SR protocol that include a detailed description of the objectives and methods that will be followed in the review and a plan for submitting the protocol for independent peer review by outside methodological and content experts. The final standard calls for making the protocol available to the public by placing it in a registry as specified by the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement (Liberati et al. 2009).

Table 4 Institute of Medicine standards for conducting systematic reviews

Standard	Description
Initiate SR	Standard 1: (*a*) Assemble team with requisite skills, (*b*) manage bias and conflict of interest, (*c*) solicit input from guideline users and stakeholders, (*d*) manage bias and conflict of interest (of users and stakeholders), (*e*) formulate the topic (analytic framework and structured PICOTS questions), (*f*) develop SR protocol for how to proceed with search, (*g*) submit protocol for peer review, and (*h*) make protocol available for public review (publish protocol).
Find and assess individual studies	Standard 2: (*a*) Conduct comprehensive, systematic search, (*b*) address potentially biased reporting of research results, (*c*) screen and select studies (two or more independent screeners), (*d*) document search, (*e*) manage data collection (two or more independent readers), and (*f*) critically appraise each study for risk of bias with respect to capacity to draw causal inference (internal validity), applicability to clinical practice (external validity), and fidelity of implementation.
Synthesize the body of evidence	Standard 3: (*a*) Use prespecified method to systematically assess the body of evidence, (*b*) conduct qualitative synthesis of available evidence as a whole, (*c*) decide whether to conduct a quantitative review (meta-analysis), and (*d*) use expert methodologists to conduct the quantitative review (if executed).
Report the SR	Standard 4: (*a*) Use structured format to prepare final report (abstract, executive summary, summary written for lay public, introduction, methods, results, and discussion), (*b*) conduct peer review with a period for public comments, and (*c*) publish final report so as to ensure public access.

Abbreviations: PICOTS, populations, interventions, comparisons, outcomes, time, and settings; SR, systematic review. From Inst. Med. (2011b).

Perhaps the most conceptually interesting standard among this initial set is formulating the topic for the SR (also known as scoping). The IOM recommends two key steps in that regard. The first is to develop an analytic framework that provides a conceptual overview of the questions of interest, including the patient and contextual factors (moderators) that might influence the outcomes of interest and the causally active treatment processes or patient mechanisms (mediators) by which those effects are produced. The analytic framework provides a conceptual blueprint that relates treatments to outcomes across relevant populations and via different means.

Figure 3 provides an example of what the analytic framework might look like with respect to depression, based on a template by Whitlock and colleagues (2002). It is provisional only; the final version to be followed will be established by the GDP working with the SR team. Our current intent is to examine treatment of unipolar depression across the life span (including child and adolescent, adult, and older adult populations), but this depends on available resources and the priorities determined by the GDP. Given that our goal is to inform the general public about the advantages and limitations of the various treatment options, the interventions of interest will include the full range of psychological, pharmacological, and somatic treatments currently applied to depression. Intermediate outcomes will include possible mechanisms of change such as psychological or biological functions. Given our interest in relational processes, it also might include measures of the therapeutic alliance or other transactional aspects of the therapeutic relationship. Clinical outcomes are likely to include symptom status, functional capacity, and quality of life.

Numbered circles in **Figure 3** identify the various types of questions that will be addressed in the review. Type 1 questions ask whether the intervention produces the desired changes in one or more of the end-state outcomes. Type 2 questions ask whether the intervention produces changes in one or more of the purported mechanisms. Type 3 questions ask whether the intervention produces changes in one or more serious adverse events, noxious side effects, or otherwise undesirable outcomes. Type 4 questions ask whether relations exist between the mediators and the final outcomes of interest.

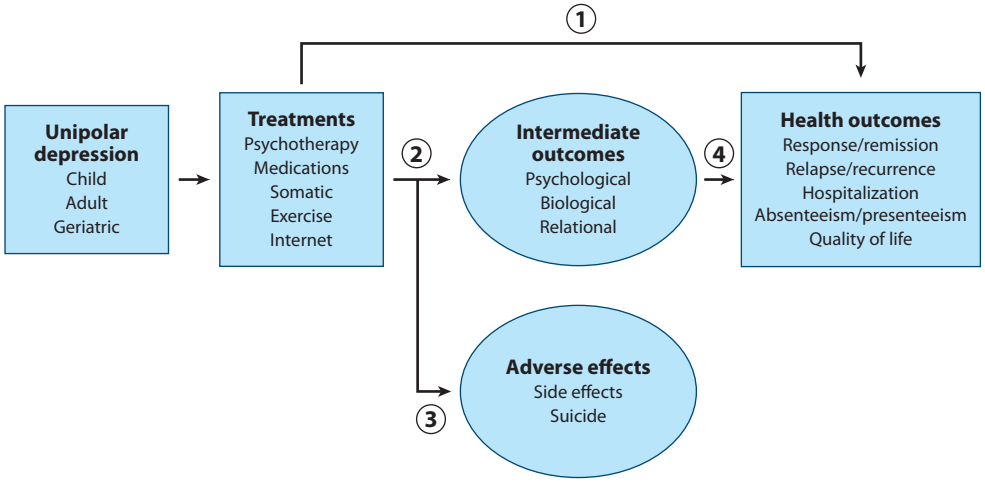


Figure 3

Example of an analytic framework. Ovals depict intermediate outcomes; rectangles depict clinical outcomes.

The second key step involved in formulating the topic is to generate structured questions that operationalize the analytic framework. We plan to adopt the widely used PICOTS format in generating these questions. PICOTS is a mnemonic that stands for populations (P), interventions (I), comparisons (C), outcomes (O), time (T), and settings (S). A specific PICOTS question might take the following form: For patients with major depressive disorder (P), is interpersonal psychotherapy (I) superior to treatment as usual (C) in terms of the reduction of acute distress (O) across the course of acute treatment (T) in outpatient settings (S)? Specific PICOTS questions can be formulated for different subtypes of depression (including but not limited to chronic, recurrent, melancholic, and atypical), different treatments (including any of the psychotherapeutic, pharmacological, and somatic interventions), different comparators (including no treatment, nonspecific controls, and alternative interventions), different outcomes (including symptomatic outcomes, functional indices, and quality of life), different temporal intervals (including acute response and short- and long-term follow-ups after initial remission or recovery), and different kinds of settings (including inpatient, outpatient, and general practice).

Organizing the information search around PICOTS questions can facilitate identification of factors that moderate or mediate response to treatment. Individuals within a larger population (P) differ in a myriad of ways, and any number of characteristics can be explored to see if they predict differential response to treatment (moderation). Therapist effects can be explored as main effects (average differences among therapists) or within the context of interventions (I) and may be associated with differential response (again an instance of moderation). Specificity can be explored via the selection of different types of comparisons (C); a treatment that is better than no treatment can be said to be efficacious, whereas a treatment that is better than generic treatment (operationalized as a nonspecific control) can be said to be specific (Chambless & Hollon 1998). Treatments are complex packages that contain both specific and nonspecific (especially relational) components, and it is by no means certain what carries the weight of the causal change. This is essentially an exercise in construct validity in that it asks whether a treatment works for reasons specified by theory (Shadish et al. 2002). Such designs provide a test of mediation and move us closer to connecting the causal steps, given that positive findings (if they exist) imply that causally active

mechanisms are being mobilized that go beyond those mobilized by simply going into treatment. Comparisons between different types of treatments can be used to identify either superiority (if one or more treatments show a consistent advantage over the others) or the absence of specificity (and the preeminence of common factors) if no reliable differences emerge (assuming studies are adequately powered to detect differences in the first place). Most treatment interventions are complex packages of theoretically specified and common factors, and the judicious selection of PICOTS questions can help to determine the extent to which each is causally active. Different kinds of outcomes (O) can be examined in different sets of questions and should be broadly defined to capture events of interest beyond symptom reduction (e.g., quality of life or functional capacity). Both temporal (T) and contextual factors (S) can be explored as moderators as well. In some instances these potential moderating factors will be represented in the designs of the individual studies and tested via treatment-by-moderator interactions (Kraemer et al. 2002), and in others instances they can be examined via the use of metaregressions in meta-analysis (Cuijpers et al. 2008). Recent conceptual and methodological advances can facilitate the exploration of such complex interventions (Hawe et al. 2004) and better understanding of the complexity of the clinical situation (Pawson et al. 2004).

Finding and assessing individual studies. The IOM recommends working with a research librarian or other information specialist to conduct a comprehensive, systematic search based on explicit inclusion and exclusion criteria to address each PICOTS question and pursue an independent peer review of the search strategy by subject matter experts outside of the SR team before proceeding. The IOM standards also address how to minimize biased reporting of research results. This includes searching the "gray" literature, such as conference abstracts and clinical trial registries, for studies that were started but never published. Biased reporting can alter the perceived efficacy of an intervention. For example, in half of the trials of selective serotonin reuptake inhibitors (SSRIs) submitted for US Food and Drug Administration approval, the drug did not outperform a pill placebo, but two-thirds of those null findings went unpublished (study publication bias) and the other third were framed so as to make the results look positive (outcome reporting bias) (Turner et al. 2008). Inclusion of the negative trials decreased the apparent efficacy of medication by a quarter. The IOM standards will limit these kinds of biases.

The IOM standards also address the process of screening and selecting studies, a labor-intensive task that is critical to the quality of the SR. The IOM recommends specifying inclusion and exclusion criteria a priori and piloting their use. It also recommends using two or more independent screeners to determine whether a given study meets those criteria. Although such independent replication is expensive, it has been shown to reduce errors in the screening process by up to a third (Edwards et al. 2002).

RCTs may be particularly well suited for establishing treatment efficacy, but they rarely have sufficient power to detect serious but uncommon events (Vandenbroucke 2004). Observational studies and patient registries play an important role in detecting the harms associated with a given intervention (Chou & Helfand 2005). Psychosocial interventions are less likely to produce serious harms and side effects than medications or somatic treatments (failure to produce benefit is more common), but untoward events do sometimes happen (Dimidjian & Hollon 2010). We plan to expand our search to include observational studies and patient registries to facilitate the detection of the kinds of negative outcomes not easily identified in RCTs (Chou et al. 2010).

The IOM recommends documenting the search with a line-by-line description of the search strategy that includes the disposition of each report identified and the reason if excluded. The IOM also recommends using at least two independent readers to extract key information from each study. Multiple reports from the same project should be linked to ensure that data from the

same study are not included more than once, and data extraction forms should be pilot tested. These recommendations will guide our decisions when commissioning SRs.

The final standard in this set calls for the SR team to critically appraise each study with respect to risk of bias using predefined criteria. For controlled experiments, we plan to use the assessment tool developed by the Cochrane Collaboration to assess possible sources of bias in randomized trials (Higgins & Altman 2008). This tool assesses for risks most directly relevant to the internal validity of the trial: (*a*) adequate generation of allocation sequence, (*b*) concealment of allocation to conditions, (*c*) prevention of knowledge of the allocated intervention (blinding), and (*d*) methods for dealing with incomplete outcome data. The IOM also recommends assessing the relevance of the studies to clinical practice (external validity) with regard to the characteristics of the target population, the nature of the interventions used (frequency, duration, and format), and the outcomes measured (what gets assessed and when). Finally, the IOM notes that it is important to assess the fidelity of treatment implementation (whether each was delivered as intended). Investigator allegiance is a powerful predictor of outcome in controlled trials (Luborsky et al. 1999), and differential competence is a major contributor to that effect (Leykin & DeRubeis 2009). Other considerations come into play when evaluating other types of designs.

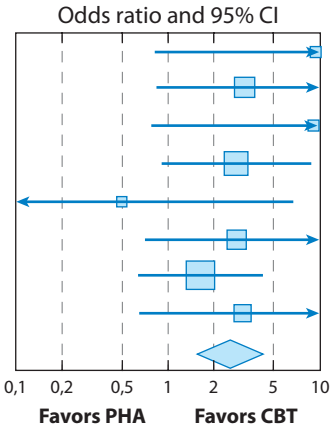
Synthesizing the body of evidence. The IOM recommends using a prespecified method to systematically assess the body of evidence. As previously indicated, we are likely to modify the GRADE system for our own use, as other societies have done. The IOM also recommends conducting a qualitative synthesis of the body of evidence as a whole that includes descriptions of the clinical and methodological characteristics of the studies (including whether important subgroups were excluded or included in insufficient numbers to yield reliable subgroup results), the strengths and limitations of the studies as a whole, flaws in the design of important studies (or groups of studies) that could bias the results as a whole, and the relevance of individual studies to the primary questions of interest. The IOM also calls for making an explicit decision regarding whether to conduct a quantitative statistical review (meta-analysis). Not all literatures contain sufficient numbers of comparable studies to justify estimating an aggregate effect. If the decision is made to go forward, the IOM recommends using expert methodologists to develop and execute the meta-analysis and conducting an independent peer review of the protocol by experts outside of the SR team before starting.

Graphic presentations are an invaluable way to represent heterogeneity across the various types of studies. **Figure 4** presents a pair of forest plots from a recently published meta-analysis comparing depression relapse rates among responders to cognitive behavioral therapy (CBT) following treatment termination versus medication responders who either terminate (panel *a*) or continue (panel *b*) medications. As can be seen in panel *a*, despite some heterogeneity (differences were not statistically significant in two of the trials), prior CBT has a relatively robust enduring effect that protects against subsequent relapse following treatment termination. Panel *b* is even more informative because it suggests a preventive effect in aggregate (a nonsignificant trend) for prior CBT relative to continuation medication, despite the fact that differences were not significant in any one of the individual studies (Cuijpers et al. 2013).

Funnel plots represent yet another kind of graphic presentation that can facilitate the interpretation of the findings from a body of evidence. In funnel plots the effect size is graphed as a function of the standard deviation or sample size. In the absence of publication bias, one would expect the distribution of study effect sizes to be normally distributed around the true mean. A dearth of studies in the lower left quadrant (small studies with weak effects) is assumed to reflect publication bias. **Figure 5** depicts just such a distribution based on studies comparing psychotherapy for depression to relevant controls (Cuijpers et al. 2010). Panel *a* depicts the asymmetric distribution

a

Study	Statistics for each study			
	Odds ratio	Lower limit	Upper limit	p-value
Blackburn (1986)	9.60	0.85	108.72	0.07
Dobson (2008)	3.25	0.88	12.01	0.08
Evans (1992)	9.00	0.81	100.14	0.07
Hollon (2005)	2.86	0.94	8.71	0.07
Jarret (2000)	0.50	0.04	6.68	0.60
Kovacs (1981)	2.88	0.73	11.38	0.13
Shea (1992)	1.66	0.65	4.21	0.29
Simons (1986)	3.15	0.67	14.86	0.15
Pooled	2.61	1.58	4.31	0.00



b

Study	Statistics for each study			
	Odds ratio	Lower limit	Upper limit	p-value
David (2008)	1.50	0.70	3.22	0.30
Dobson (2008)	2.80	0.88	8.91	0.08
Evans (1992)	2.00	0.15	26.19	0.60
Hollon (2005)	1.64	0.59	4.57	0.34
Jarret (2000)	0.25	0.02	2.76	0.26
Pooled	1.62	0.97	2.72	0.07

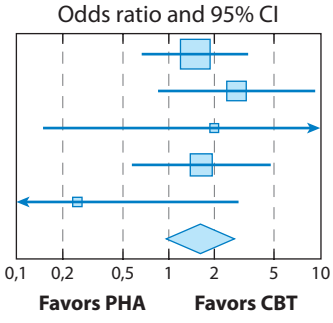


Figure 4

Example of a forest plot. (*a*) Long-term effects of cognitive-behavioral therapy (CBT) (without continuation during follow-up) compared with pharmacotherapy (PHA) (discontinued during follow-up): forest plot of odds ratio of response. (*b*) Long-term effects of CBT (without continuation during follow-up) compared with pharmacotherapy (continued during follow-up): forest plot of odds ratio of response. Reprinted with permission from Cuijpers et al. (2013).

of psychotherapy studies, whereas panel *b* depicts how many studies would have to be imputed to make the distribution symmetrical. This provides an estimate of the number of studies presumed missing because of nonpublication of null findings from small trials. Imputing studies to correct for publication bias decreases the estimated effect size for psychotherapy by about a quarter, similar to what was found when unpublished trials were included in the medication literature (Turner et al. 2008).

Reporting the systematic review. The final set of IOM standards concern reporting the systematic review. The IOM recommends using a structured format that includes an abstract, executive summary, and summary written for the lay public. It also recommends including an introduction

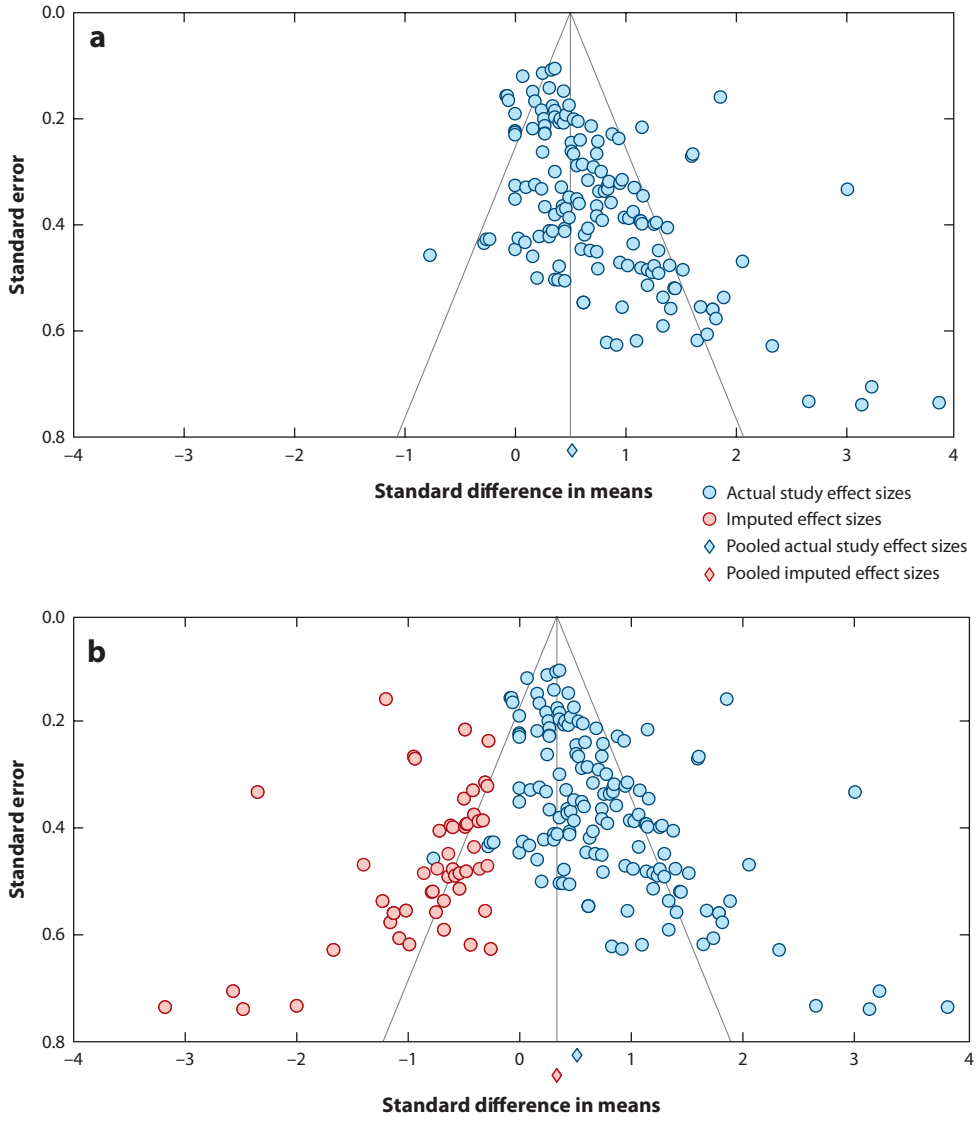


Figure 5

Funnel plot of the mean effect size and standard error. (*a*) Unadjusted for publication bias. (*b*) Adjusted for publication bias. Reprinted with permission from Cuijpers et al. (2010).

that describes the rationale and objectives of the SR; a method section that describes the research protocol, including the analytic framework, key PICOTS questions, and methods to appraise study quality; a results section that is organized around the key PICOTS questions, appraises individual study quality, and provides the results of the meta-analysis (if applicable) and tables and figures; and a discussion section that includes a summary of the evidence, strengths and limitations of the SR, conclusions for each key question, gaps in the evidence, and future research needs. The IOM also calls for conducting a peer review in conjunction with inviting public comment before publishing the final report to ensure public access.

We intend to follow each of these standards and the associated recommendations to the best of our ability (given that our resources are limited), unless otherwise indicated. The goal is to provide a comprehensive and unbiased summary of the available evidence that can be presented to an independent GDP with varying perspectives for their consideration. The GDP will be charged with weighing the relevant evidence through the prism of clinical judgment to generate recommendations regarding treatment options useful to patients, providers, and decision makers.

KEY CONSIDERATIONS

Pros and Cons of Clinical Practice Guidelines

Some worry that CPGs are mandates that override clinician judgment and client preferences. As noted in the introduction, that is not the case. CPGs are recommendations based on the best available evidence as filtered through clinical judgment (IOM 2008). While aiming to reduce variation in the quality of care by recommending the best care possible, they are not intended to replace the judgment of the independent clinician (Wolf et al. 2011). In this section, we review the advantages and limitations of CPGs.

Advantages. The most noteworthy advantage is that patients are more likely to receive care that is informed by the best, most up-to-date evidence available, which in turn leads to improved outcomes. CPGs inform clinicians as to which interventions have the most demonstrated benefit and which are not supported by the evidence. They also call attention to potentially harmful or wasteful interventions. SRs, upon which the CPGs are based, provide an invaluable synthesis of the literature for busy clinicians who may not have the time to keep up-to-date with the relevant research (Bastian et al. 2010). A number of rigorously conducted medical studies have shown that implementation of CPGs can improve quality of care (e.g., Grimshaw et al. 2012) and health outcomes (Wolf et al. 2011). Similar findings are emerging from CPGs for psychological conditions such as anxiety (van Dijk et al. 2012).

A second advantage is that guidelines can reduce unjustified variation in what care is delivered. CPGs have the potential to inform health care policies and ensure that psychological interventions are appropriately covered and thus available to everyone despite disparities in income and locale. Additionally, government agencies often rely on CPGs to inform their decisions as to the services made available to underserved populations. In the absence of guidelines, these agencies may turn to other, more biased sources of information. By calling attention to under-recognized health problems, neglected patient populations, and high risk groups, the CPGs can change public policy to promote service delivery to those conditions and groups of the population most in need (Woolf et al. 1999), although it is best if coverage policy is not directly linked to the guidelines themselves (Woolf & Campos-Outcalt 2013).

A third benefit is the potential for CPGs to reduce confusion among patients who seek care for a mental health or behavioral problem. With contradictory statements about best practices widely available on the Internet, CPGs endorsed and created by the APA can provide credible and consistent information to patients about the best treatments for health conditions. As such, patients can be empowered "to make informed health care choices and to consider their personal needs and preferences in selecting the best option" (Woolf et al. 1999, p. 527).

A final important benefit is that guidelines provide a road map for future research by highlighting areas in which evidence is lacking in quantity or quality. They provide a foundation from which the research community can build, in terms of particular problems, populations, or settings that are in need of more empirical attention.

Limitations. Relevant scientific evidence may be lacking, misleading, or misinterpreted, and the composition of the guideline panel or COIs may bias the interpretations and recommendations generated (Kung et al. 2012, Shaneyfelt 2012, Wolf et al. 2011, Woolf et al. 1999). To mitigate this limitation, we have chosen to follow the IOM standards reviewed in prior sections. However, even with these methodological standards in place, there will be occasions when the data are lacking or uncertain, leaving the GDP to make recommendations based on clinical consensus (clearly identified as such) or to make no recommendations at all (Moodie et al. 2011). We will aim to reduce this outcome by selecting topics for which there exists an adequate evidence base for recommendations. Nonetheless, "even when the data are certain, recommendations for or against interventions will involve subjective value judgments when the benefits are weighed against the harms" (Woolf et al. 1999, p. 529).

A second potential limitation is that if the guidelines are inflexible or limited in their scope, they "can harm by leaving insufficient room for clinicians to tailor care to patients' personal circumstances and medical history" (Woolf et al. 1999, p. 529). In other words, the guidelines come from average behaviors of "acceptably similar" groups of patients, and thus they may not be appropriate for every individual (Kent et al. 2010). Addressing important subgroups of the population in the guidelines can mitigate this risk. More importantly, we propose to follow the IOM/APA premise that guidelines are to guide but not supplant clinician judgment. CPGs need not be rigid, but they are essential because they overcome the limitations and biases inherent in sole reliance on clinical judgment, including selective use of evidence, motivation, memory distortions, professional norms, and business pressures (Moodie et al. 2011).

A third potential limitation is that CPGs could lead to problematic public policy changes, such as denial of insurance coverage for interventions that are not recommended or that are discouraged, even though the patient and clinician believe such interventions have been effective. Similarly, funding priorities may be shifted away from interventions that are not recommended and toward recommended ones, thus potentially negatively affecting important new discoveries. In this case, CPGs need to be clear about which interventions are truly ineffective, which are promising but need more study, and which established interventions need to be improved. Absence of evidence is not evidence of absence, and drawing attention to interventions that are in need of further testing can offset these negative impacts. A related concern is that the guidelines could place restraints on the development of experimental or novel treatments. Because no treatment is universally efficacious this should be resisted.

A fourth issue regarding the impact of CPGs is that there may be significant barriers to their implementation. One barrier is the quality of the guideline itself. We will follow the Appraisal of Guidelines Research and Evaluation (AGREE) instrument in developing guidelines that are comprehensive and valid (AGREE Collab. 2003, Brouwers et al. 2010). Other barriers are the time and resources required to learn new skill sets, and simple resistance to change (Moodie et al. 2011). CPGs need to be accompanied by training programs and systems of care that invest in the value of implementation. Without such support, practice is unlikely to change.

A final consideration is that the users of CPGs may overemphasize adherence at the cost of competency. Adherence refers to the degree to which therapists follow a prescribed approach without deviation, whereas competence refers to the skill with which they do so with respect to both the technical and relational aspects of the therapy (Shaw et al. 1999). Adherence and competence are distinct but related concepts. For therapists to be competent with a given treatment, they must be adherent to that intervention, but adherence to the intervention provides no guarantee of competence (Waltz et al. 1993). CPGs should emphasize that whereas adherence is important, competency is even more so, and that the ability to deliver an intervention flexibly and in an interpersonally skillful fashion can be critical.

Specific and Nonspecific Treatment Components

There has been a long-running controversy in health care as to whether specific treatment strategies have causal effects over and above those produced by nonspecific therapeutic factors. Too often the discussion has been cast in "either/or" terms: Change is viewed as a consequence of either nonspecific or specific factors, although treatments are actually a complex interaction between specific and nonspecific factors, as well as therapist and client characteristics. Nonspecific factors are always present (to a greater or lesser extent), and specific factors add to the effects of the nonspecifics. Nonspecific factors include positive expectancy, instillation of hope, and relationship features (alliance, empathy, etc.). Nonspecific factors likely contribute to treatment of all disorders and may be sufficient for some (like mild depression) but not others (like severe depression) (Driessen et al. 2010, Fournier et al. 2010). Specific factors range from single therapeutic activities to a series of therapeutic techniques delivered in a prescribed manner to comprehensive treatment programs, all with theoretically derived purported mechanisms of change that extend beyond nonspecific factors (Magnavita 2005, 2010; Sexton et al. 2011).

Whenever possible, CPGs will evaluate the evidence from SRs on therapeutic effects as a function of the nonspecific factors common to all treatments and as a function of the specific mechanism(s) mobilized by a given treatment, with each influenced by patient, therapist, and setting factors. Because of the inherent complexity of these elements and the interactions among them, it is useful to have an analytic framework that specifies the various components to guide the search for relevant information in the SR and to provide a working framework for the various GDPs.

Underrepresented Populations and Gaps in the Literature

RCTs offer foundational evidence in CPGs because they permit the most valid causal inferences about whether different treatments work. Inevitably, though, CPG developers grapple with several kinds of gaps in the research literature. First, some novel, potentially promising treatments have been understudied and not evaluated in well-conducted RCTs. Second, some subsets of the population have been omitted from trials of well-established treatments, raising a question about whether the efficacy of those treatments can be assumed to apply to all sectors of the population. Third, the evidence base on treatment effectiveness has been focused largely on negative outcomes, such as symptom reduction or the elimination of problematic behaviors. Some would say that positive outcomes that are highly valued by much of the population, including good quality of life, well-being, happiness, and family harmony, have been understudied as treatment targets (Seligman et al. 2006).

Racial and ethnic minorities continue to be underrepresented in clinical trials (Eap & Hall 2008; Hoel et al. 2009; IOM 2001; Miranda et al. 2003, 2005; Rochon et al. 2004). This is particularly poignant given the health disparities they experience, as described in the Surgeon General's report on the impact of culture, race, and ethnicity in mental health (US Dep. Health Hum. Serv. 2001). As a simple matter of fairness, there is a need for research to more equitably support the quality of care for all, especially those who have borne the impact of disparate treatment and outcomes. This gap is made even more problematic by the fact that the United States is moving in the direction of becoming a so-called majority-minority country; by 2050 whites, the current racial majority, will make up less than half of the population. Clearly, CPGs need to be able to speak to the current landscape of available health care options and the needs of diverse racial and ethnic groups. The ability to do so requires having an available evidence base that is both sufficiently inclusive of current treatments and also representative of the populations targeted for care. CPGs can highlight the gaps in patient-centered outcome research that must be addressed to generate evidence-based recommendations.

Engaging diverse stakeholders in our GDPs is one vehicle through which we aim to identify and highlight understudied treatments and outcomes. As a promising development, needed relevant evidence may emerge from research now being supported by the Patient-Centered Outcomes Research Institute (PCORI) established by the 2010 Patient Protection and Affordable Care Act (Pub. L. 111-148). All PCORI research must be patient centered, aiming to learn what patients can do to improve the outcomes that are most important to them. Patients, community members, and other stakeholders participate as co-leaders in all PCORI research, from determining which treatment decisions most urgently need an evidence base to choosing which outcomes are most meaningful to particular groups. This has been shown to enhance identification of which evidence gaps need to be filled to let guidelines better serve the needs, concerns, and priorities of diverse, real-world populations.

Developmental Issues

Far too often, CPGs for mental and behavioral disorders have focused only on young to middle-aged adults, rather than on the special developmental issues related to the prevention, treatment, and management of mental and behavioral disorders in children, adolescents, and older adults. The population in the United States is aging considerably—the baby boom generation is now entering late life. As a result, clinicians will wonder whether the treatments they provide to their young to middle-aged only adult population are applicable to older adults and what adaptations may be needed to make existing treatments more appropriate for older adults. Children and adolescents are ideal targets for prevention. When children and adolescents suffer from mental illness or exhibit behavioral problems, clinicians are faced with important questions, such as whether and when the potential positive impact of an intervention outweighs the stigmatizing effect of singling out the child to receive treatment.

How children, adolescents, and older adults enter the mental health system is markedly different from how young to middle-aged adults seek care. Children rarely come into treatment of their own accord (Ollendick & Shirk 2011). More typically, they are brought into treatment by their parents, who believe their child has a problem. In contrast, when adults come into treatment, it is because they themselves believe they have a problem. A barrier that deters older adults from seeking treatment is the belief that mental health difficulties are a normal consequence of aging. Therefore, older adults rarely seek treatment because they expect to become depressed or feel anxious (Areán et al. 2002). These basic differences in presentation set the stage for important developmental differences to emerge and may warrant special PICOTS questions to support work with children, adolescents, and older adults.

Developmental influences are known to moderate responsiveness to certain types of treatment. For example, younger children benefit most from simple, task-oriented self-instruction, whereas older children and adolescents benefit more from generalized self-instruction and from instructions that combine specific, task-oriented statements and general strategy statements (Ollendick & Shirk 2011). Older adults with major depression and comorbid deficits in executive functions (e.g., planning, information and emotional processing, and problem solving) respond poorly to antidepressant medication (Alexopoulos et al. 2002a,b) but respond exceptionally well to psychotherapies, in particular, psychotherapies that target executive skills (e.g., problem-solving therapy) (Alexopoulos et al. 2011, Areán et al. 2010).

Finally, optimal methods for treating mental health and behavioral problems in children, adolescents, and older adults often involve team-based care. Effective intervention for many behavioral issues among children and adolescents requires the involvement of family and school systems in treatment planning and symptom management. For older adults, the medical and social work

systems are important in treatment planning, meaning that team-based treatments are the most effective for conditions such as late-life depression (Unützer et al. 2002).

These developmental issues illustrate the need for CPGs to go beyond simply matching treatments to disorders and to instead incorporate such contextual factors as age, culture, education, and environment. For these reasons, we have enlisted individuals who provide a developmental and contextual expertise to serve on the various GDPs. To the extent feasible, we plan to take a life span perspective on each of the phenomena we address.

Promoting Guideline Implementation

Clinical practice guidelines are intended to address the well-established research-to-practice translational gap that is reflected by the fact that much of the public fails to receive effective, warranted health care (McGlynn et al. 2003). Dissemination of clinical practice guidelines to practitioners and consumers is a necessary, but not sufficient, step to ensure their use (Giguère et al. 2012). Whereas dissemination involves methods to bring guidelines to the attention of therapists and consumers, implementation entails strategies to promote the integration of research evidence into health care practice, systems, and policy.

Certain features increase a guideline's odds of being implemented. For instance, Burgers and colleagues (2005) determined that the likelihood of successful implementation is increased if a guideline has (*a*) concrete aims and objectives, (*b*) sufficient evidence to support most recommendations, (*c*) a clear structure and attractive layout, (*d*) clear and specific recommendations, and (*e*) applicability in different settings and (*f*) takes into account the norms and values of the target users (p. 71). Statements that readily translate into actual behaviors make it easier to determine if the end user followed the guideline recommendation. For example, "Medication should be switched if the patient has not responded after four weeks at maximum tolerated dose" is an actionable recommendation, whereas "Medication should be switched as needed" is not. Increasingly, guidelines are being made actionable by being programmed into technologies, including electronic health records, Internet programs, and smartphone applications in the form of decision-making algorithms (Shojania et al. 2010). These decision support tools guide the clinician through a typical sequence of treatment decisions and suggest responses (e.g., "If the patient has not responded to first-line treatment A, consider second-line treatments X, Y, and Z."). The clinician can accept the recommendation or not depending on whether it accords with clinical judgment and patient preferences.

Whereas the dissemination of guidelines and accompanying educational materials has some modest positive effect on professional practice, the impact on implementation is greater when accompanied by multifaceted and tailored interventions that address specific barriers to change (Baker et al. 2010, Grimshaw et al. 2012). In addition to numerous contextual factors to be considered, familiar barriers include client expectations, lack of training and confidence about delivering new forms of treatment, and competing local norms (Damschroder et al. 2009). Helpful strategies include the provision of needed education and support and the involvement of thought leaders who can facilitate adoption by others (McHugh & Barlow 2010).

A national guidelines implementation plan would be consistent with existing APA policies bearing on education and practice but is not currently being prepared. APA more likely will develop a suite of tools to be adapted for local implementation of CPGs. Some prominent examples are the online tools developed by NICE (2012) that include proposed metrics to monitor implementation of key guideline recommendations. Also, like both NICE and the APA Education Directorate, the APA will identify staff contacts to develop resources to assist with implementation, such as designated contacts for consultation as well as online training modules or web-based forums to

discuss implementation issues. Finally, as part of guideline development, attention to products suitable for incorporation into electronic health records or other health information technology will be a key aspect of successful implementation.

FUTURE DIRECTIONS

Making Better Treatment Choices

An important function of CPGs is to highlight gaps in what is currently known. There are a number of widely practiced treatment modalities that have not been adequately tested. From the perspective of the general public, it is important to know whether those interventions actually work and, if so, how they compare to other viable alternatives. Advocates for those approaches would do well to remember that although absence of evidence is not necessarily evidence of absence (of effect), it sometimes is considered justification for absence of reimbursement. CPGs also can help in selecting the best treatment for a given patient. Patient characteristics sometimes predict differential response to different treatments. If these moderators replicate across studies (or can be detected through metaregressions), they can be used to guide treatment selection so as to increase the likelihood of response for an individual patient.

Making Better Treatments

CPGs also help identify when existing treatments are simply not sufficient. Treatments should at least be better than their absence (efficacious), and it is good to document whether they rise above generic treatment (specificity) or other alternative interventions (superiority). In some sense, anything short of full normalization and the elimination of future risk represents a failure of existing interventions. Yet despite tremendous strides in treatment research over the last 50 years, disorders like the schizophrenias or autism can neither be prevented nor even adequately treated. Even for an eminently treatable disorder like depression, only about a third of all treated patients show full remission and survive even a year without relapse (Hollon 2011).

Moreover, the fact that a treatment is effective does not mean that it is as effective as it could be, nor perhaps as efficient as it may need to be to warrant patient adherence or third-party reimbursement. New research methods adapted from the engineering sciences are now being applied widely to optimize treatments to any desired criterion, including effect size, reach into the population, efficiency, cost, patient characteristics, or treatment history (Collins et al. 2011, Lei et al. 2012). These methodological advances illustrate the type of leap forward that can result when scientists from different disciplines pool their respective expertise. The current burgeoning of transdisciplinary science and interprofessional practice reflects a growing appreciation that diversity can be a source of strength and that the clash of paradigms that results when people with different perspectives work on common problems can spur important innovations (Spring et al. 2012).

Clinical innovation is as likely to come from the consulting room as from the laboratory. Novel ideas are at least as likely to come from anecdotal experience (the context of discovery) as from a careful program of research (the context of verification), and the latter is better suited to testing existing theories than to generating them. We lose something if we ignore the insights provided by individual clinicians struggling to help their patients. That is where some of the most powerful ideas have come from in the history of modern clinical science.

Ongoing Assessment, Evaluation, and Quality Improvement

Treatments work best when they are implemented with fidelity and ongoing feedback about client and therapist performance (Sexton et al. 2011). Attaining a good clinical outcome is even more

dependent on treatment fidelity when an intervention is delivered by interventionists in the community rather than by research staff (McHugh et al. 2009). In addition to benefiting from feedback about their own performance, clinicians practice more effectively when ongoing outcome assessments keep them abreast of how their clients are doing (Lambert 2010). Measurement feedback systems have been designed to provide timely feedback about both intervention implementation and client responses, creating the basis for integrated tools that support continual quality improvement. The creation of such digitally enabled systems is in keeping with the IOM's (2012) aspiration to build a digitally enabled learning health care system based on continual knowledge development, improvement, and application. Such a system can be mined to cull real-time insights from routine patient care, creating evidence that may well be incorporated into future CPGs.

Making Better Guidelines

Finally, we strongly endorse the notion of joining with other health care disciplines like psychiatry, social work, and nursing to develop joint CPGs. We fully endorse the IOM's call for interdisciplinary representation on any given GDP and strongly encourage the relevant professional organizations to work together to generate future guidelines. To supplement the government's efforts in guideline generation, it falls to the various professional organizations to set aside guild concerns and work together in the public interest. We also strongly support the efforts of organizations like the Guideline International Network to pool resources across nations in the effort to improve the quality of clinical services on a truly global basis.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

AGREE Collab. 2003. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual. Saf. Health Care* 12(1):18–23

Alexopoulos GS, Kiosses DN, Choi SJ, Murphy CF, Lim KO. 2002a. Frontal white matter microstructure and treatment response of late-life depression: a preliminary study. *Am. J. Psychiatry* 159(11):1929–32

Alexopoulos GS, Kiosses DN, Klimstra S, Kalayam B, Bruce ML. 2002b. Clinical presentation of the "depression-executive dysfunction syndrome" of late life. *Am. J. Geriatr. Psychiatry* 10(1):98–106

Alexopoulos GS, Raue PJ, Kiosses DN, Mackin RS, Kanellopoulos D, et al. 2011. Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction: effect on disability. *Arch. Gen. Psychiatry* 68(1):33–41

Am. Psychol. Assoc. (APA). 1995. *Template for Developing Guidelines: Interventions for Mental Disorders and Psychosocial Aspects of Physical Disorders*. Washington, DC: Am. Psychol. Assoc.

Am. Psychol. Assoc. (APA). 2002. Criteria for evaluating treatment guidelines. *Am. Psychol.* 57:1052–59

APA Pres. Task Force Evid. Based Pract. 2006. Evidence-based practice in psychology. *Am. Psychol.* 61(4):271–85

Areán PA, Alvidrez J, Barrera A, Robinson GS, Hicks S. 2002. Would older medical patients use psychological services? *Gerontologist* 42(3):392–98

Areán PA, Raue P, Mackin RS, Kanellopoulos D, McCulloch C, Alexopoulos GS. 2010. Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction. *Am. J. Psychiatry* 167(11):1391–98

Baker R, Camosso-Stefanovic J, Gilliss CL, Shaw EJ, Cheater F, et al. 2010. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. *Cochrane Database Syst. Rev.* 17(3):CD005470

Bastian H, Glasziou P, Chalmers I. 2010. Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Med.* 7(9):e1000326

Beutler LE, Forrester B, Thompson-Gallagher D, Thompson L, Tomlins JB. 2012. Common, specific, and treatment fit variables in psychotherapy outcome. *J. Psychother. Integr.* 22(3):255–81

Blackburn IM, Eunson KM, Bishop S. 1986. A two-year naturalistic follow-up of depressed patients treated with cognitive therapy, pharmacotherapy and a combination of both. *J. Affect. Disord.* 10:67–75

Borgerson K. 2009. Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspect. Biol. Med.* 52(2):218–33

Brouwers M, Kho ME, Browman GP, Burgers JS, Cluzeau F, et al. 2010. AGREE II: advancing guideline development, reporting and evaluation in healthcare. *Can. Med. Assoc. J.* 182:E839–42

Burgers JS, Grol R, Eccles M. 2005. Clinical guidelines as a tool for implementing change in patient care. In *Improving Patient Care: The Implementation of Change in Clinical Practice*, ed. R Grol, M Wensing, M Eccles, pp. 71–92. Edinburgh, UK: Elsevier

Chambless DL, Hollon SD. 1998. Defining empirically supported therapies. *J. Consult. Clin. Psychol.* 66(1):7–18

Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, et al. 2010. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health care program. *J. Clin. Epidemiol.* 63(5):502–12

Chou R, Helfand M. 2005. Challenges in systematic reviews that assess treatment harms. *Ann. Intern. Med.* 142(12 Pt. 2):1090–99

Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, et al. 2011. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann. Behav. Med.* 41(2):208–26

Cuijpers P, Hollon SD, van Straten A, Bockting C, Berking M, Andersson G. 2013. Does cognitive behavior therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? *BMJ Open* 3(4):e002542

Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. 2010. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br. J. Psychiatry* 196(3):173–78

Cuijpers P, van Straten A, van Oppen P, Andersson G. 2008. Are psychological and pharmacological interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J. Clin. Psychiatry* 69(11):1675–85

Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. 2009. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 7:4–50

David D, Szentagotai A, Lupu V, Cosman D. 2008. Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: a randomized clinical trial, posttreatment outcomes, and six-month follow-up. *J. Clin. Psychol.* 64:728–46

Dimidjian S, Hollon SD. 2010. How would you know if psychotherapy was harmful? *Am. Psychol.* 65(1):21–33

Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, et al. 2008. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *J. Consult. Clin. Psychol.* 76:468–77

Driessen E, Cuijpers P, Hollon SD, Dekker JJM. 2010. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J. Consult. Clin. Psychol.* 78(5):668–80

Eap S, Hall GCN. 2008. Relevance of RCTs to diverse groups. In *Evidence-Based Outcome Research: A Practical Guide to Conducting Randomized Clinical Trials for Psychosocial Interventions*, ed. AM Nezu, CM Nezu, pp. 425–43. New York: Oxford Univ. Press

Edwards P, Clark M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. 2002. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Stat. Med.* 21(11):1635–40

Evans MD, Hollon SD, DeRubeis RJ, Piasecki J, Grove WM, et al. 1992. Differential relapse following cognitive therapy and pharmacotherapy for depression. *Arch. Gen. Psychiatry* 49:802–8

Falck-Ytter Y, Schünemann HJ. 2009. *Rating the evidence: using GRADE to develop clinical practice guidelines*. Presented at Agency Healthc. Res. Qual. Annu. Conf., Sep. 13–16, Bethesda, MD. **http://www.ahrq.gov/news/events/conference/2009/falck-ytter-schunemann/index.html**

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, et al. 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *J. Am. Med. Assoc.* 303(1):47–53

Giguère A, Légaré F, Grimshaw J, Turcotte S, Fiander M, et al. 2012. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst. Rev.* CD004398

Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squire JE. 2012. Knowledge translation of research findings. *Implement. Sci.* 7:50

Guyatt G, Gutterman D, Bauman MH, Addrizzo-Harris D, Hylek EM, et al. 2006. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians task force. *Chest* 129(1):174–81

Harris RP, Hefland M, Woolf SH, Lohr KN, Mulrow CD, et al. 2001. Current methods of the US Preventive Services Task Force: a review of the process. *Am. J. Prevent. Med.* 20(3 Suppl.):21–35

Hawe P, Shiell A, Riley T. 2004. Complex interventions: How "out of control" can a randomized controlled trial be? *BMJ* 328(7455):1561–63

Higgins JPT, Altman DG. 2008. Chapter 8: assessing risk of bias in included studies. In *Cochrane Handbook for Systematic Reviews of Interventions*, ed. JPT Higgins, S Green. Version 5.0.1. **http://www.cochrane-handbook.org**

Hoel AW, Kayssi A, Brahmanandam A, Belkin M, Conte MS, Nguyen LL. 2009. Under-representation of women and ethnic minorities in vascular surgery in randomized controlled trials. *J. Vasc. Surg.* 50(2):349–54

Hollon SD. 2011. Cognitive and behavior therapy in the treatment and prevention of depression. *Depress. Anxiety* 28(4):263–66

Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, et al. 2005. Prevention of relapse following cognitive therapy versus medications in moderate to severe depression. *Arch. Gen. Psychiatry* 62:417–22

Inst. Med. (IOM). 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: Natl. Acad. Press

Inst. Med. (IOM). 2008. *Knowing What Works in Health Care: A Roadmap for the Nation*. Washington, DC: Natl. Acad. Press

Inst. Med. (IOM). 2011a. *Clinical Practice Guidelines We Can Trust*. Washington, DC: Natl. Acad. Press

Inst. Med. (IOM). 2011b. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: Natl. Acad. Press

Inst. Med. (IOM). 2012. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: Natl. Acad. Press

Jarrett RB, Kraft D, Schaffer M, Witt-Browder A, Risser R, et al. 2000. Reducing relapse in depressed outpatients with atypical features: a pilot study. *Psychother. Psychosom.* 69:232–39

Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. 2010. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 11:85

Kovacs M, Rush AJ, Beck AT, Hollon SD. 1981. Depressed outpatients treated with cognitive therapy or pharmacotherapy: a one-year follow-up. *Arch. Gen. Psychiatry* 38:33–39

Kraemer HC, Wilson GT, Fairburn CG, Agras WS. 2002. Mediators and moderators of treatment effects in randomized clinical trials. *Arch. Gen. Psychiatry* 59(10):877–83

Kung J, Miller RR, Mackowiak PA. 2012. Failure of clinical practice guidelines to meet Institute of Medicine standards: two more decades of little, if any, progress. *Arch. Intern. Med.* 172(21):1628–33

Lambert MA. 2010. *Prevention of Treatment Failure: The Use of Measuring, Monitoring, and Feedback in Clinical Practice*. Washington, DC: Am. Psychol. Assoc.

Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. 2012. A "SMART" design for building individualized treatment sequences. *Annu. Rev. Clin. Psychol.* 8(21):21–48

Leykin Y, DeRubeis RJ. 2009. Allegiance in psychotherapy outcome research: separating association from bias. *Clin. Psychol. Sci. Pract.* 16(1):54–65

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, et al. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann. Intern. Med.* 151(4):W65–94

Luborsky L, Diguer L, Seligman DA, Rosenthal R, Krause ED, et al. 1999. The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. *Clin. Psychol. Sci. Pract.* 6(1):95–106

Magnavita JJ. 2005. *Personality-Guided Relational Psychotherapy: A Unified Approach*. Washington, DC: Am. Psychol. Assoc.

Magnavita JJ, ed. 2010. *Evidence-Based Treatment of Personality Dysfunction: Principles, Methods, and Processes*. Washington, DC: Am. Psychol. Assoc.

Maier SF, Amat J, Baratta MV, Paul E, Watkins LR. 2006. Behavioral control, the medial prefrontal cortex, and resilience. *Dialogues Clin. Neurosci.* 8(4):397–406

Marcus SC, Olfson M. 2010. National trends in the treatment of depression from 1998 to 2007. *Arch. Gen. Psychiatry* 67(12):1265–73

McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, et al. 2003. The quality of health care delivered to adults in the United States. *N. Engl. J. Med.* 348(26):2635–45

McHugh RK, Barlow DH. 2010. The dissemination and implementation of evidence-based psychological treatments: a review of current events. *Am. Psychol.* 65(2):73–84

McHugh RK, Murray HW, Barlow DH. 2009. Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: the promise of transdiagnostic interventions. *Behav. Res. Ther.* 47(11):946–53

Mellers B, Hertwig R, Kahneman D. 2001. Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychol. Sci.* 12(4):269–75

Minami T, Wampold BE, Serlin RC, Kircher JC, Brown GS. 2007. Benchmarks for psychotherapy efficacy in adult major depression. *J. Consult. Clin. Psychol.* 75(2):232–43

Miranda J, Bernal G, Lau A, Kohn L, Hwang WC, LaFromboise T. 2005. State of the science on psychosocial interventions for ethnic minorities. *Annu. Rev. Clin. Psychol.* 1:113–42

Miranda J, Nakamura R, Bernal G. 2003. Including ethnic minorities in mental health intervention research: a practical approach to a long-standing problem. *Cult. Med. Psychiatry* 27(4):467–86

Moodie ST, Kothari A, Bagatto MP, Seewald R, Miller LT, Scollie SD. 2011. Knowledge translation in audiology: promoting the clinical application of best evidence. *Trends Amplif.* 15(1):5–22

Natl. Inst. Health Care. Excell. (NICE). 2012. *Implementation Tools*. London: NICE. **http://www.nice.org.uk/usingguidance/implementationtools/implementation_tools.jsp**

Newberry SJ, Ahmadzai N, Motala A, Tsertsvadze A, Maglione M, et al. 2013. *Surveillance and Identification of Signals for Updating Systematic Review: Implementation and Early Experience*. Methods Res. Rep. AHRQ Publ. No. 13-EHC088-EF. Rockville, MD: Agency Healthc. Res. Qual.

Olfson M, Marcus SC, Druss B, Elinson L, Tanielian T, Pincus HA. 2002. National trends in the outpatient treatment of depression. *J. Am. Med. Assoc.* 287(2):203–9

Ollendick TH, Shirk SR. 2011. Clinical interventions with children and adolescents: current status, future directions. In *Oxford Handbook of Clinical Psychology*, ed. DH Barlow, pp. 762–88. New York: Oxford Univ. Press

Pawson R, Greenhalgh T, Harvey G, Walshe K. 2004. *Realist synthesis: an introduction*. ESRC Res. Methods Progr. Work. Pap. 2, Univ. Manchester

Rochon PA, Mashari A, Cohen A, Misra A, Laxer D, et al. 2004. The inclusion of minority groups in clinical trials: problems of under representation and under reporting of data. *Account. Res.* 11(3–4):215–23

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. 1996. Evidence-based medicine: what it is and what it isn't. *BMJ* 312(7023):71–72

Schünemann HJ, Brożek J, Oxman AD. 2009. *GRADE Handbook for Grading the Quality of Evidence and the Strength of Recommendations*. Grade Work. Group, version 3.2, updated Mar. **http://www.who.int/hiv/topics/mtct/grade_handbook.pdf**

Seligman MEP, Rashid T, Parks AC. 2006. Positive psychology. *Am. Psychol.* 61:772–88

Sexton T, Gordon GC, Gurman A, Lebow J, Holtzworth-Munroe A, Johnson S. 2011. Guidelines for classifying evidence-based treatments in couple and family therapy. *Fam. Process.* 50(3):377–92

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin

Shaneyfelt T. 2012. In guidelines we cannot trust: comment on "Failure of clinical practice guidelines to meet Institute of Medicine standards." *Arch. Intern. Med.* 172(21):1633–34

Shaw BF, Elkin I, Yamaguchi J, Olmsted M, Vallis TM, et al. 1999. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *J. Consult. Clin. Psychol.* 67(6):837–46

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, et al. 1992. Course of depressive symptoms over follow-up: findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Arch. Gen. Psychiatry* 49:782–87

Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. 2010. Effect of point-of-care computer reminders on physician behaviour: a systematic review. *Can. Med. Assoc. J.* 182(5):E216–25

Simons AD, Murphy GE, Levine JE, Wetzel RD. 1986. Cognitive therapy and pharmacotherapy for depression: Sustained improvement over one year. *Arch. Gen. Psychiatry* 43:43–49

Spring B, Hitchcock K. 2009. Evidence-based practice in psychology. In *Corsini's Encyclopedia of Psychology*, ed. IB Weiner, WE Craighead, pp. 603–7. New York: Wiley. 4th ed.

Spring B, Moller AC, Falk-Krzesinski HJ, Hall K, eds. 2012. Special Section on Team Approaches to Science, Practice, and Policy in Health. *Transl. Behav. Med.* Vol. 2(4). New York: Springer

Stricker G, Abrahamson DJ, Bologna NC, Hollon SD, Robinson EA, Reed GM. 1999. Treatment guidelines: the good, the bad, and the ugly. *Psychotherapy* 36:69–79

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. *New Eng. J. Med.* 358(3):252–60

Unützer J, Katon W, Callahan CM, Williams JW Jr, Hunkeler E, et al. 2002. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *J. Am. Med. Assoc.* 288(22):2836–45

US Census Bur. 2008. An older and more diverse nation by midcentury. Washington, DC: US Census Bur. **http://www.census.gov/newsroom/releases/archives/population/cb08-123.html**

US Dep. Health Hum. Serv. 2001. *Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General.* Rockville, MD: Subst. Abus. Ment. Health Serv. Adm. **http://www.ncbi.nlm.nih.gov/books/NBK44243/**

van Dijk MK, Oosterbaan DB, Verbraak MJ, van Balkom AJ. 2012. The effectiveness of adhering to clinical-practice guidelines for anxiety disorders in secondary mental health care: the results of a cohort study in the Netherlands. *J. Eval. Clin. Pract.* 79(5):791–97

Vandenbroucke JP. 2004. Benefits and harms of drug treatments: Observational studies and randomised trials should learn from each other. *BMJ* 329(7456):2–3

Vandenbroucke JP. 2008. Observational research, randomized trials, and two views of medical science. *PLoS Med.* 5(3):e67

Waltz J, Addis ME, Koerner K, Jacobson N. 1993. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *J. Consult. Clin. Psychol.* 61(4):620–630

Westen D, Novotny CM, Thompson-Brenner H. 2004. Empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychol. Bull.* 130(4):631–63

Whitlock EP, Orleans CT, Pender N, Allan J. 2002. Evaluating primary care behavioral counseling interventions: an evidence-based approach. *Am. J. Prev. Med.* 22(4):267–84

Wolf JS Jr, Hubbard H, Faraday MM, Forrest JB. 2011. Clinical practice guidelines to inform evidence-based clinical practice. *World J. Urol.* 29(3):303–9

Woolf SH, Campos-Outcalt D. 2013. Severing the link between coverage policy and the US Preventive Services Task Force. *J. Am. Med. Assoc.* 309(18):1899–900

Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. 1999. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 318(7182):527–30