

Summary¹

*Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?*
T.S. Eliot, *The Rock* (1934)

In the early 21st century, despite unprecedented advances in biomedical knowledge and the highest per capita health care expenditures in the world, the quality and outcomes of health care vary dramatically across the United States. The economic burden of health spending is weakening American industry's competitive edge and consumers are increasingly asked to take on a greater share of the burden. Consumer-directed health care is viewed by some as a means to rationalize what most agree is a health system plagued by overuse, underuse, and misuse. Yet even the most sophisticated health consumer struggles to learn which care is appropriate for his or her circumstance.

It is in this context that the Robert Wood Johnson Foundation asked the Institute of Medicine (IOM) to examine how the nation uses scientific evidence to identify highly effective clinical services. The IOM appointed the Committee on Reviewing Evidence to Identify Highly Effective Clinical Services in June 2006 to respond to the foundation's request (Box S-1). The committee was charged with recommending a sustainable, replicable approach to identifying effective clinical services. Ultimately, the committee concluded that the nation must significantly expand its capacity to use scientific evidence to assess "what works" in health care. This report recommends an organizational framework for a national clinical effectiveness assessment program, referred to throughout as "the Program." The Program's mission would be to optimize the use of evidence to identify effective health services. Three functions would be central to this mission: setting priorities for evidence assessment, assessing evidence (systematic review), and developing (or endorsing) standards for trusted clinical practice guidelines.

BOX S-1 Charge to the IOM Committee

The committee was charged with recommending a sustainable, replicable approach to identifying and evaluating the clinical services that have the highest potential effectiveness. The charge specified three principal tasks:

- (1) To recommend an approach to identifying highly effective clinical services across the full spectrum of health care services—from prevention, diagnosis, treatment, and rehabilitation, to end of life care and palliation.
- (2) To recommend a process to evaluate and report on evidence on clinical effectiveness.
- (3) To recommend an organizational framework for using evidence reports to develop recommendations on appropriate clinical applications for specified populations.

¹ This summary does not include references. Citations for the findings presented in the summary appear in the subsequent chapters.

CONCEPTUAL FRAMEWORK

The committee based its work on the central premise that decisions about the care of individual patients should be based on the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. This means that individual clinical expertise should be integrated with the best information from scientifically-based, systematic research and applied in light of the patient's values and circumstances. Centering decision-making on the patient is integral to improving the quality of health care and is also imperative if consumers are to take an active role in making informed health care decisions based on known risks and benefits. This report also recognizes that health care resources are finite. Thus, setting priorities for systematic assessment of scientific evidence is essential.

The era of physician as sole health care decision maker is long past. In today's world, health care decisions are made by multiple people, individually or in collaboration, in multiple contexts for multiple purposes. The decision maker is likely to be the consumer choosing among health plans, patients or patients' caregivers making treatment choices, payers or employers making health coverage and reimbursement decisions, professional medical societies developing practice guidelines or clinical recommendations, regulatory agencies assessing new drugs or devices, or public programs developing population-based health interventions. Every decision maker needs credible, unbiased, and understandable evidence on the effectiveness of health interventions and services.

What constitutes evidence that a health service is effective? Scientists view evidence of effectiveness as knowledge that is explicit, systematic, and replicable. However, patients, clinicians, payers, and other decision makers, often have a different, more contextual perspective on what constitutes evidence of effectiveness. Decision makers consider the scientific evidence as demonstrating what works under ideal circumstances, but of necessity are also interested in "real world" circumstances. Patient factors such as comorbidities, underlying risk, adherence to therapies, disease stage and severity, health insurance coverage, and demographics; intervention factors such as care setting, level of training, and timing and quality of intervention; and other factors can affect the applicability of the results of an individual study to a particular clinical decision or circumstance. There cannot be a single study that covers all populations, intervention approaches, and settings related to a clinical question. Systematic reviews of multiple high quality studies have the advantage of providing summaries of the available research, which typically covers many different circumstances, and providing a snapshot of where more research is needed.

The conceptual context for this study is the continuum that begins with research evidence, then moves to systematic review of the overall body of evidence, and then to the interpretation of the strength of the overall evidence for developing credible, clinical practice guidelines (Figure S-1). Individual studies rarely provide definitive answers to clinical effectiveness questions. A "systematic review" is a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize similar but separate studies. Systematic reviews are critical to developing agendas for further research because they reveal where evidence is insufficient and additional research is needed. Moreover, a systematic review of studies on clinical effectiveness provides an essential bridge between the body of research evidence and the development of clinical guidance.

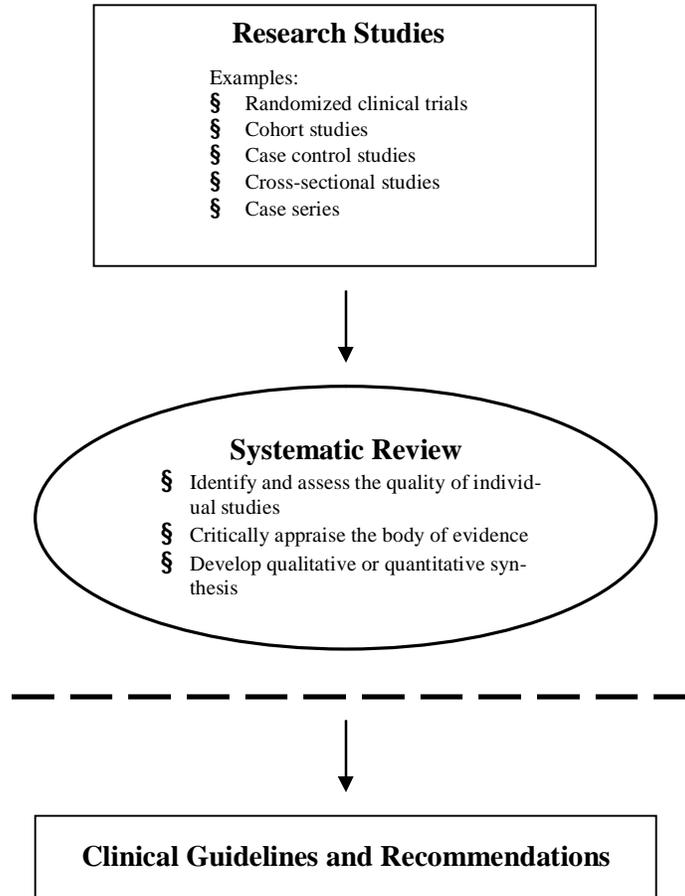


FIGURE S-1 Continuum from research studies to systematic review to development of clinical guidelines and recommendations.

NOTE: The dashed line is the theoretical dividing line between the systematic review of the research literature and its application to clinical decision making, including the development of clinical guidelines and recommendations. Below the dashed line, decision makers and developers of clinical recommendations interpret the findings of systematic reviews to decide which patients, health care settings, or other circumstances they relate to.

SOURCE: Adapted from West, S., V. King, T. Carey, K. Lohr, N. McCoy, S. Sutton, and L. Lux. 2002. *Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment No. 47. (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016.* Rockville, MD: Agency for Healthcare Research and Quality.

AN IMPERATIVE FOR CHANGE

The committee believes that unbiased, reliable information about what works in health care is essential to addressing several persistent health policy challenges (described below).

- *Constraining health care costs.* A significant proportion of health care costs are directed to care that has not been shown to be effective and may actually be harmful.
- *Reducing geographic variation in the use of health care services.* Variations in treatment patterns often reflect deviations from accepted care standards or uncertainty and

disagreement regarding what those standards should be. Uncertainties about what works and for whom means patients cannot always be assured that they will receive the best, most effective care.

- *Improving quality.* To promote quality health care, scientific knowledge should be employed, but the evidence base needed to support effective care is in many instances lacking.
- *Consumer-directed health care.* Many policy makers believe in empowering consumers and patients to be prudent managers of their own health and health care. However, consumers need information on the effectiveness, risks, and benefits of alternative treatments if they are to search for and obtain high-value treatments. The current dearth of such information is a substantial obstacle to consumer empowerment.
- *Making health coverage decisions.* Private and public health plans are struggling with an almost daily challenge of learning how their covered populations might benefit—or be harmed by—newly available health services.

LIMITATIONS IN THE STATUS QUO

There is ample evidence that, under the status quo, there are critical limitations in how the United States identifies and uses evidence on clinical effectiveness, particularly with respect to three interrelated processes: (1) setting priorities for evidence assessment; (2) assessing evidence through systematic reviews; and (3) developing trusted clinical practice guidelines.

Setting Priorities for Evidence Assessment

If we are to resolve current deficiencies in how the nation uses scientific evidence to identify the most effective clinical services, there must be a process for identifying the most important topics in order to preserve resources for evidence assessment itself. Most health technology assessment programs have an organized process for determining which topics merit comprehensive study. But currently no one agency or organization in the United States assumes a broad, national perspective on new as well as established health interventions across all populations—children as well as elderly persons, women as well as men, and including ethnic and racial minorities.

The basic elements of a priority setting process include: identifying potential topics; selecting the priority criteria, reducing the initial list of nominated topics to a smaller set to be pursued; and choosing the final priority topics. Some approaches also incorporate quantitative methods that involve collecting data to weigh priorities, assigning scores for each criterion to each topic, and calculating priority scores for each topic to produce a ranked priority list. The process is typically conducted by a committee or advisory group which reviews and chooses the topics that will be funded. It may employ a formal method, such as the Delphi Technique, to systematically develop the high priority list.

The committee could not find any systematic assessments of the comparative strengths and weaknesses of different approaches to setting priorities, including whether complex, quantitative and resource intensive methods are more effective than less rigorous approaches. Many organizations report using the same general criteria to gauge the potential impact that an evidence assessment might have on clinical care and patient outcomes. These include burden of disease (rates of disability, morbidity, or mortality), public controversy, cost (related to the condition, the procedure, in the aggregate), new evidence that might change previously held conclusions (new

clinical trial results), adequacy of the existing evidence, and unexplained variation in use of services. How these factors play into final priorities is not apparent.

At present, there is substantial unnecessary duplication in reviews of new and emerging technologies. Decision makers, especially in health plans and health systems, often need to learn quickly about new and emerging technologies and what is known and not known about effectiveness. Patients and providers want information on new health services as soon as they become available, often because manufacturers are pressing them to adopt a product or because consumers have been exposed to direct-to-consumer advertising and want answers from their physician. Yet, almost by definition, sufficient objective information about new and emerging technologies is seldom available. New and emerging technologies may require a different priority setting process—including separate criteria—than other topics with more substantive evidence.

Systematic Reviews are the Central Link Between Evidence and Clinical Decision Making

Systematic reviews of evidence on the effectiveness of health services provide a central link between the generation of research and clinical decision making. Individual studies rarely provide definitive answers to clinical effectiveness questions. If conducted properly, the systematic review should make obvious the gap between what is known about the effectiveness of a particular service and what clinicians and patients want to know. As such, systematic reviews are also critical to developing the agenda for further primary research because they reveal where evidence is insufficient and new information is needed. Without systematic reviews, researchers may miss promising leads or pursue questions that have been answered already.

Systematic review is itself a science, a new and dynamic science with evolving methods. In medicine, early implementers were trialists who saw the need to summarize data from multiple effectiveness trials, many of them with very small samples. By the late 1980s, systematic reviews were increasingly used to assess the effectiveness of health interventions but research also began to reveal problems in their execution. The methods underlying the reviews were often neither objective nor transparent. The approach to deciding which literature to include and which findings to present was subjective and nonsystematic. Still today, the quality of published reviews is variable and often unreliable.

The core of a systematic review is a concise and transparent synthesis of the results of the included studies. The language of the review should be simple and clear so that it is usable and accessible to decision makers. The synthesis may be purely qualitative, that is, describing study results individually but not combined, or it may be complemented by meta-analysis that combines the individual study results quantitatively and allows statistical inference.

Under the status quo, judging the quality of reviews is often difficult because methods are so poorly documented. Reviews rely on many disparate grading schemes and evidence hierarchies which are often not well understood. Since the underlying rationale for hierarchies is to present study designs in terms of increasing protections against bias, evidence hierarchies have the potential to raise awareness that some forms of evidence are more trustworthy than others. However, hierarchies are often oversimplified and consider just the type of research (e.g., a clinical trial versus an observational study) and not the question being asked. Observational and experimental studies each can provide valid and reliable evidence, but their relative value depends on the clinical question. For example, RCTs can best answer questions about the efficacy of screening, preventive, and therapeutic interventions while observational studies are generally the most

appropriate for answering questions related to prognosis, diagnostic accuracy, incidence, prevalence, and etiology.

The synthesis should collate, describe, and summarize the following key features of the individual studies it reviews that could have a bearing on the findings:

- Characteristics of the patient population, care setting, and type of provider.
- Intervention (route, dose, timing, duration)
- Comparison group
- Outcome measures and timing of assessments
- Quality of the evidence (i.e., risk of bias) from individual studies and possible influence on findings. The term “bias” has different meanings depending on the context in which it is used. It may refer to “bias” due to conflicts of interest. “Bias” also refers to statistical bias, i.e., the tendency for a study to produce results that depart systematically from the truth. Statistical biases can lead to under- or over-estimation of the effectiveness of an intervention.
- Sample sizes
- Quantitative results and analyses including examination of whether the study estimates of effect are consistent across studies.
- Examination of potential sources of study heterogeneity if relevant.

The synthesis should not include recommendations. If the systematic review is both scientific and transparent, decision makers should be able to interpret the evidence, to know what is not known, and to describe the extent to which the evidence is applicable to clinical practice and particular subgroups of patients. Making evidence-based decisions—such as when a guideline developer recommends what should and should not be done in specific clinical circumstances—is a distinct and separate process from systematic review.

It is not known how many researchers in the United States are adequately trained and qualified to conduct systematic reviews on the effectiveness of health services.

Developing Evidence-Based Clinical Practice Guidelines

The development of clinical guidelines in the United States today is highly decentralized and involves many public and private organizations—medical professional societies, patient advocacy groups, payers, government agencies, and others. The National Guideline Clearinghouse (NGC) maintained by the Agency for Healthcare Research and Quality includes clinical guidelines from about 360 different organizations. The U.S. Preventive Services Task Force produces recommendations for preventive services that are widely considered to offer a gold standard for the process of guideline development. International organizations also produce clinical guidelines that are available in the United States.

One of the challenges inherent in having a highly decentralized, pluralistic process for developing clinical guidelines is that multiple groups will produce guidelines in the same clinical topic area. Currently, for example, the NGC contains 471 guidelines relating to the topic of hypertension and 276 guidelines related to stroke. Despite the abundance of clinical guidance for some topics, there is little clinical guidance on other important topics.

The translation of evidence into recommendations is not straightforward. Although guideline developers have adopted several strategies to improve the reliability and trustworthiness of the information they provide, it is not yet possible to say that the development of clinical guidelines

is based on a scientifically validated process. The key challenges stem from the fact that guideline development frequently forces organizations to go beyond available evidence to make practical recommendations for use in everyday practice. Given the gaps in the evidence base that frequently exist and the variable quality of the information that is available, some observers have suggested that one criterion of an effective guideline process is to have two separate grading systems: one for the quality of evidence and another for the recommendations themselves. Even when there is substantial consensus about the existing scientific evidence, there may be different interpretations about what the evidence means for clinical practice. Different interpretations can be due, for example, to conflicting viewpoints about which outcomes are the most important or which course of action is appropriate given that evidence is imperfect.

RECOMMENDATIONS

The committee recommends the development of a national clinical effectiveness assessment program to facilitate the development of standards and processes that yield credible, unbiased, and understandable syntheses of the available evidence on clinical effectiveness for patients, individual clinicians, health plans, purchasers, specialty societies, and others. The committee hopes that the nation now has the will to address the urgent need to bolster the U.S. health system with a foundation built on research evidence and scientific methods.

The committee recommends a single entity be established to help determine what works in health care. Box S-2 lists all the recommendations presented in this report. Each recommendation is elaborated on in its respective chapter with a rationale and strategies for implementation.

Recommendation: Congress should direct the secretary of the U.S. Department of Health and Human Services to designate a single entity (the Program) with authority, overarching responsibility, sustained resources, and adequate capacity to ensure production of credible, unbiased information about what is known and not known about clinical effectiveness. The Program should:

- **Set priorities for, fund, and manage systematic reviews of clinical effectiveness and related topics.**
- **Develop a common language and standards for conducting systematic reviews of the evidence and for generating clinical guidelines and recommendations.**
- **Provide a forum for addressing conflicting guidelines and recommendations.**
- **Prepare an annual report to Congress.**

The committee further recommends that an advisory board be appointed to oversee the Program, and that the Program develop (or endorse) standards to minimize bias.

Recommendation: The secretary of Health and Human Services should appoint a Clinical Effectiveness Advisory Board to oversee the Program. Its membership should be constituted to minimize bias due to conflict of interest and should include representation of diverse public and private sector expertise and interests.

Recommendation: The Program should develop standards to minimize bias due to conflicts of interest for priority setting, evidence assessment, and recommendations development.

The committee envisions a Program—whether a public entity or a public-private entity—that develops standards and sets priorities and facilitates systematic reviews of priority topics by external organizations. The committee believes that the most pragmatic—and also the most promising—approach to establishing such a Program is to build on current efforts. In addition, private organizations that currently produce guidelines, such as professional societies and others, treasure their autonomy and would likely oppose efforts to reduce their role. Further, guidelines that have the imprimatur of a respected professional society are able to engender trust in end users. Finally, there are some indications that the quality of these guidelines has improved over time.

BOX S-2 Recommendations

Building a Foundation for Knowing What Works in Health Care (Chapter 6)

Congress should direct the secretary of the U.S. Department of Health and Human Services to designate a single entity (the Program) with authority, overarching responsibility, sustained resources, and adequate capacity to ensure production of credible, unbiased information about what is known and not known about clinical effectiveness. The Program should:

- § Set priorities for, fund, and manage systematic reviews of clinical effectiveness and related topics.
- § Develop a common language and standards for conducting systematic reviews of the evidence and for generating clinical guidelines and recommendations.
- § Provide a forum for addressing conflicting guidelines and recommendations.
- § Prepare an annual report to Congress.

The secretary of Health and Human Services should appoint a Clinical Effectiveness Advisory Board to oversee the Program. Its membership should be constituted to minimize bias due to conflict of interest and should include representation of diverse public and private sector expertise and interests.

The Program should develop standards to minimize bias due to conflicts of interest for priority setting, evidence assessment, and recommendations development.

Setting Priorities (Chapter 3)

The Program should appoint a standing Priority Setting Advisory Committee (PSAC) to identify high priority topics for systematic reviews of clinical effectiveness.

- § The priority setting process should be open, transparent, efficient, and timely.
- § Priorities should reflect the potential for evidence-based practice to improve health outcomes across the life span, reduce the burden of disease and health disparities, and eliminate undesirable variation.
- § Priorities should also consider economic factors, such as the costs of treatment and the economic burden of disease.
- § The membership of the PSAC should include a broad mix of expertise and interests and be chosen to minimize committee bias due to conflicts of interest.

Assessing Evidence (Chapter 4)

The Program should develop evidence-based methodologic standards for systematic reviews, including a common language for characterizing the strength of evidence. The Program should fund reviewers only if they commit to and consistently meet these standards.

- § The Program should invest in advancing the scientific methods underlying the conduct of systematic reviews and, when appropriate, update the standards for the reviews it funds.

The Program should assess the capacity of the research workforce to meet the Program's needs, and, if deemed appropriate, it should expand training opportunities in systematic review and comparative effectiveness research methods.

Developing Clinical Practice Guidelines (Chapter 5)

Groups developing clinical guidelines or recommendations should use the Program's standards, document their adherence to the standards, and make this documentation publicly available.

To minimize bias due to conflicts of interest, panels should include a balance of competing interests and diverse stakeholders, publish conflict of interest disclosures, and prohibit voting by members with material conflicts.

Providers, public and private payers, purchasers, accrediting organizations, performance measurement groups, patients, consumers, and others should preferentially use clinical recommendations developed according to the Program standards.

The committee wants to ensure that the national Program recommended by the committee is stable over the long term; its output is judged as objective, credible, and without conflict of interest or bias; and its operations are independent of external political pressures. For that reason, the committee recommends that the Program be built on the basis of eight core principles: accountability, consistency, efficiency, feasibility, objectivity, responsiveness, scientific rigor, and transparency (Box S-3).

| BOX S-3: Program Principles | |
|------------------------------------|--|
| Accountability | Parties are directly responsible for meeting standards. |
| Consistency | Processes are predictable and standardized so as to be readily usable by patients, health professionals, medical societies, payers, and purchasers. |
| Efficiency | Avoids waste and unnecessary duplication. |
| Feasibility | Capable of operating in the real world; recognizing political, economic, and social implications. |
| Objectivity | Evidence-based and without bias, e.g., balanced participation, governance, and standards minimize conflicts of interest and other biases. |
| Responsiveness | Addresses information needs of decision makers in a timely way. Able to react quickly. Patients and health professionals require real time information for treatment decisions. |
| Scientific rigor | Methods minimize bias, provide reproducible results, and are completely reported. |
| Transparency | Methods are explicitly defined, consistently applied, and available for public review so that observers can readily link judgments, decisions, or actions to the data on which they are based. |

Recommendations for Setting National Priorities for Systematic Reviews

Setting national priorities for systematic reviews is important because the overall value of the Program will hinge, in part, on how effectively the enterprise determines its priorities. The committee recommends that the Program appoint an independent, free-standing Priority Setting Advisory Committee—the PSAC—to develop and implement a priority setting process that will identify those high priority topics which merit systematic evidence assessment. In contrast to the Clinical Effectiveness Advisory Board which should provide broad oversight of the Program, the PSAC should be an active advisory body that meets frequently to advise the Program on topics that merit priority systematic review.

Recommendation: The Program should appoint a standing Priority Setting Advisory Committee (PSAC) to identify high priority topics for systematic reviews of clinical effectiveness.

- **The priority setting process should be open, transparent, efficient, and timely.**

- **Priorities should reflect the potential for evidence-based practice to improve health outcomes across the life span, reduce the burden of disease and health disparities, and eliminate undesirable variation.**
- **Priorities should also consider economic factors, such as the costs of treatment and the economic burden of disease.**
- **The membership of the PSAC should include a broad mix of expertise and interests and be chosen to minimize committee bias due to conflicts of interest.**

The PSAC should consider a broad range of topics, including, for example, new, emerging, and well-established health services across the full spectrum of health care (e.g., preventive interventions, diagnostic tests, treatments, rehabilitative therapies, palliation, and end-of-life care); community-based interventions such as immunization initiatives or programs to encourage smoking cessation; and research methods and data sources for the analysis of comparative effectiveness.

The highest priorities should focus on the clinical questions of patients and clinicians that have the potential for substantial impact on health outcomes across all ages, burden of disease and health disparities, and undesirable variation in the delivery of health services.

There is limited research evidence to suggest the optimal composition or size of the PSAC. The committee believes it should be sufficiently large to include all of the important stakeholders, but not too large so that it is unwieldy. The membership should mirror the Program's target audience, especially patients and consumers, clinicians, payers, purchasers, and guideline developers, and also individuals with the appropriate expertise in relevant content areas and technical methods.

The PSAC should cast a wide net to include all stakeholders in an open and transparent topic nomination process. The process should especially cultivate input from end users such as guideline developers, consumers, patients, health professionals, and payers. While the nomination process should not be overly burdensome to potential nominators, there should be standardized methods and information requirements.

Objectivity implies balanced participation, oversight by a governance body, and standards that minimize conflicts of interest and other biases.² The PSAC should not be dominated by special interests that can benefit materially or by intellectual biases that might favor one professional specialty over another (e.g., surgery versus medicine, ophthalmology versus optometry).

Using transparent, well-documented and standard procedures also contribute to perceptions of objectivity. Stakeholders are not likely to trust an unpredictable, opaque process. All deliberations should be open to encourage public participation, public confidence, and ensure a wide variety of perspectives. The PSAC should post key documents on its website, including meeting announcements and decisions concerning priorities, and give time for public comment on documents that support the priority setting process.

² IOM has recently appointed the Committee on Conflict of Interest in Medical Research, Education, and Practice to recommend principles for managing conflicts of interest in the conduct of medical research, development of practice guidelines, and patient care. A final report is expected in 2009 and may provide important guidance to the Program.

Recommendations for Conducting Systematic Reviews

Recommendation: The Program should develop evidence-based, methodologic standards for systematic reviews, including a common language for characterizing the strength of evidence. The Program should fund reviewers only if they commit to and consistently meet these standards.

- **The Program should invest in advancing the scientific methods underlying the conduct of systematic reviews and, when appropriate, update the standards for the reviews it funds.**

Recommendation: The Program should assess the capacity of the research workforce to meet the Program's needs, and, if deemed appropriate, it should expand training opportunities in systematic review and comparative effectiveness research methods.

Recommendations for Developing Trusted Clinical Practice Guidelines

Clinical practice guidelines vary widely in their methodological rigor and protection from bias and the committee recommends that steps be taken to ensure that the information communicated through practice guidelines is trustworthy.

Recommendation: Groups developing clinical guidelines or recommendations should use the Program's standards, document their adherence to the standards, and make this documentation publicly available.

Recommendation: To minimize bias due to conflicts of interest, panels should include a balance of competing interests and diverse stakeholders, publish conflict of interest disclosures, and prohibit voting by members with material conflicts.

Recommendation: Providers, public and private payers, purchasers, accrediting organizations, performance measurement groups, patients, consumers, and others should preferentially use clinical recommendations developed according to the Program standards.