6

Building a Foundation for Knowing What Works in Health Care

Abstract: *The committee recommends that Congress direct the secretary of the U.S. Department of Health and Human Services to establish a single national clinical effectiveness assessment program (the Program) with the authority and resources to set priorities for and sponsor systematic reviews of clinical effectiveness, and to develop methodologic and reporting standards for conducting systematic reviews and developing clinical guidelines. The secretary should appoint a broadly representative Clinical Effectiveness Advisory Board to oversee the Program. This chapter considers three alternative approaches to building the Program infrastructure: the status quo, a central agency model, and a hybrid model. In the previous chapters, the committee found convincing evidence that systematic reviews and clinical guidelines are often of poor quality, lacking scientific rigor and objectivity. The committee observed that, under the status quo, systematic reviews and clinical guidelines are produced by numerous public and private organizations with little or no coordination, minimal quality controls, inconsistent terminology, inadequate transparency, and without concerted attention to the priorities of all types of consumers, patients, and other stakeholders. The committee finds that a highly centralized Program, such as in a central agency, the quality of both evidence assessment and guideline development may be tightly controlled. But such an agency would be costly and take too much time to establish. Thus, the committee recommends that the Secretary build on existing capacity to establish the Program infrastructure (the hybrid approach), with substantial stakeholder involvement and strict standards to protect against bias and conflict of interest.*

The United States must substantially strengthen its capacity for scientific inquiry into evidence on what is known and not known about what works in health care. Under the status quo, there is not enough objective and credible information identifying which health services work best, for whom, and under what circumstances (Medicare Payment Advisory Commission, 2007). Interest in a national comparative clinical effectiveness program is growing. Recently, the Medicare Payment Advisory Commission concluded unanimously that because information on clinical effectiveness can benefit all users and is a public good, the federal government should act to produce unbiased information and make it publicly available (Medicare Payment Advisory Commission, 2007). Other stakeholders and analysts agree (America's Health Insurance Plans, 2007; BCBSA, 2007b; Congressional Budget Office, 2007; IOM, 2007; Kupersmith et al., 2005; Shortell et al., 2007; Wilensky, 2006).

The previous chapters examined three essential functions—priority setting, evidence assessment (systematic review), and developing clinical practice guidelines—of a national clinical effectiveness assessment Program ("the Program). This chapter explores how best to approach establishing an infrastructure for organizing the three functions. It first reviews the foundational principles that the committee adopted to guide its analysis and then assesses three alternatives

(i.e., the status quo, a central agency model, and a hybrid model). The chapter concludes with the committee's recommendations regarding the program infrastructure.

GUIDING PRINCIPLES

During the course of this study, a number of important themes emerged that led the committee to establish a set of guiding principles for building the Program. These themes include convincing evidence (described in the previous chapters) that financial and other types of conflicts of interest may compromise the integrity of research findings and related clinical recommendations, indications that a meaningful proportion of evidence reviews frequently lack scientific rigor, and current efforts fall far short of addressing patients' and health professionals' need for current, trustworthy information on clinical effectiveness. The committee particularly wants to ensure that its recommended Program will be stable over the long term, that its output be judged as objective and meeting broadly accepted standards of scientific rigor, that it will be useful to stakeholders, that it is without conflict of interest or bias,[1] and that its operations be independent of external political pressures.

In developing and defining its guiding principles, the committee also drew from important foundational work performed by others—most notably, several earlier Institute of Medicine (IOM) committees, including the Committee on Quality of Health Care in America, the Committee on Setting Priorities for Guidelines Development, and the Committee on Priorities for Assessment and Reassessment of Health Care Technologies; the Agency for Healthcare Research and Quality (AHRQ); the Cochrane Collaboration; the AGREE Collaboration (Appraisal of Guidelines Research and Evaluation); the GRADE Working Group; and the National Quality Forum (AHRQ, 2007; AGREE Collaboration, 2001; Cochrane Collaboration, 2007; GRADE Working Group, 2004; IOM, 1992, 1995, 2001; NQF, 2006).

Box 6-1 defines eight guiding principles for organizing the Program: accountability, consistency, efficiency, feasibility, objectivity, responsiveness, scientific rigor, and transparency. The committee believes that each principle is integral to ensuring a valued, effective enterprise that instills credibility and trust in its products. The following sections further describe each principle.

[1] The term "bias" has different meaning depending on the context in which it is used. Here it refers to "bias" due to conflicts of interest. In discussions regarding systematic review methods, "bias" refers to statistical bias, i.e., the tendency for a study to produce results that systematically depart from the truth.

BOX 6-1 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, up-to-date 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.

Accountability

For the Program, accountability refers to accepting the responsibility to meet and demonstrate compliance with a set of program performance standards. Under the status quo, a meaningful proportion of systematic reviews of clinical effectiveness are proprietary and their findings are available only to those who pay for them. The documentation on the methods used to conduct systematic reviews is uneven and often lacking, even when the review and analysis are presented in a journal or some other public medium (Moher et al., 2007). As a result, it may be impossible to determine if the review process was free from bias and met scientific and performance standards.

Consistency

Consistency refers to the use of standardized and predictable methods. It is an important element not only in a program's regulations and administrative procedures, but also in its analytic methods and products. Although a number of organizations and individuals currently generate high-quality evidence syntheses, potential users of the information are often frustrated by unexplained differences in the terminologies, methods, and conclusions. When reviews present methods and findings in a uniform way, it is easier for the user to appraise the evidence as a whole and assess the underlying differences in the findings from studies assessing a similar question. Another advantage of consistency is that it makes it easier for manufacturers to make accurate predictions of budgets for the evaluation of new technologies and new applications of existing technologies for product evaluation.

Efficiency

Efficiency means the avoidance of waste and the effective use of resources. Setting national priorities for which services should be evaluated can help avoid unnecessary duplication and can also focus limited resources on the most important questions. It is not efficient for every payer, provider organization, or medical professional society to invest in assessment of the same topics. Guideline developers and payers faced with coverage decisions are overburdened with duplicating production of systematic reviews. Numerous private sector organizations, such as health plans and technology assessment firms, set their own priorities for assessing evidence but their research is often duplicative as many parties tend to focus on the same set of emerging technologies and new applications of existing technologies (BCBSA, 2007a; ECRI, 2006; Hayes, 2006). While some duplication may be desirable and private organizations should be free to set their own research priorities, users of evidence have little basis for deciding which available reviews to rely upon.

Feasibility

For a program to be feasible it must be able to function in the real world; its processes must be sound, its resources must be adequate over the long term, and its leaders must pay attention to stakeholders. A program must also be attuned to political realities. If the program lacks sufficient public support, it will be neither implemented nor sustained. If the program is not protected from political conflict and funding is withdrawn, the public investment will be wasted and any gains made will be lost. This lesson has been repeated numerous times during the decades of on-and-off federal involvement in research on clinical effectiveness (Congressional Budget Office, 2007). In particular, the committee notes the experience of AHRQ as an example of political pressures that have short-circuited the important beginnings of high-quality clinical effectiveness research in the United States. In the early 1990s, funding for AHRQ was almost eliminated due to stakeholders' anger over the findings presented in its guideline on interventions for back pain (Gray, 1992; Gray et al., 2003).

Objectivity

Objectivity requires the incorporation of certain features in a program, such as balanced participation, governance, and standards that minimize conflicts of interest and other biases. Objectivity is central to the development of public confidence in the integrity of an organization. Patients, health professionals, payers, and developers of practice guidelines depend on systematic reviews to know whether the available evidence is valid. They need to be able to trust the program to reach conclusions that are driven solely by the evidence and never by special interests that may benefit materially. The public will not trust a program that does not have adequate protections against bias and conflict of interest.

As the previous chapters have described, there is a growing literature documenting that in comparison with non-industry-sponsored research, industry-sponsored research—including evidence reviews—is more likely to favor the sponsor's product (Lexchin et al., 2003). Financial interests are not the only source of bias. Program participants may have intellectual biases (e.g., regarding their own body of work), or program processes may favor one professional specialty over another (e.g., surgery versus medicine, ophthalmology versus optometry).

Although it may not always be possible to make a process entirely free from bias, there are always steps that can be taken to address areas of concern. For example, many studies of devices

and drugs are funded by their manufacturers. Given legitimate concerns about reporting biases, detailed information about funding sources should always be made public. Moreover, systematic reviews should indicate the funding source not only for the individual studies, but also for the review itself. The Program may find advice from a forthcoming report from the IOM Committee on Conflict of Interest in Medical Research, Education, and Practice. The committee is developing guidance for managing conflicts of interest in the development of clinical practice guidelines and conduct of medical research. A final report is expected in 2009.

Responsiveness

The overall value of the Program will hinge, in part, on how responsive it is to the information needs of decision makers, i.e., patients, clinicians, health plans, purchasers, specialty societies, and other decision makers. No mechanism currently insures that evidence assessments address the concerns of all types of patients or all types of services across the continuum of care. In many cases, evidence on effectiveness does not extend to children, older individuals, minority populations, people with multiple conditions, or particular community settings; and new research may be warranted (National Research Council, 2004; Simpson, 2004).

Responsiveness also implies timeliness including an obligation to stay current on the topics of research. The frequency with which reviews need updating depends on the production of valid new evidence. The Cochrane Collaboration recommends that systematic reviews be updated every 2 years or should have a commentary to explain why this is done less frequently. This recommendation has been supported by a recent study conducted by Shojania and colleagues (2007). The investigators analyzed the need for updates of 100 clinically relevant systematic reviews of drugs, devices, and procedures that signaled the need for an update, such as new trial evidence reversing the findings of an earlier effectiveness review. They found that almost one in four reviews (23 percent) needed an update within 2 years of publication of the reviews, 15 percent within 1 year, and 7 percent before publication.

Scientific Rigor

As applied to evidence reports and recommendation statements, scientific rigor implies that research methods minimize bias, that the results are reliable and valid, and that both the methods used and all results are completely reported. Methods have been developed for systematically reviewing evidence on effectiveness and these methods are evidence based (i.e., the evidence has shown that failure to adhere to these methods can result in invalid or biased findings) (Higgins and Green, 2006; Moher et al., 1999; Stroup et al., 2000). However, as noted earlier, there is considerable evidence indicating that many systematic reviews do not meet scientific standards (Gøtzsche et al., 2007; Moher et al., 2007). Particularly worrisome is the lack of attention to the quality and scientific rigor of the studies that are included in the review. Publication in a high-impact journal, unfortunately, does not guarantee that the methods used in the study were sound (Steinberg and Luce, 2005). Less is known about bias-free processes for translating evidence into clinical recommendations.

Transparency

In the present context, transparency refers to the use of clear, unambiguous language to convey scientific results and conclusions. It gives the reader the ability to clearly link judgments, decisions, or actions to the information on which they are based. Different entities frequently review the same published evidence and arrive at different conclusions about their safety and ef-

fectiveness, and it is important to be able to identify possible explanations. Methods should be 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. There is extensive evidence that most systematic reviews lack adherence to a transparent and documented set of standards (Bhandari et al., 2001; Delaney et al., 2005; Glenny et al., 2003; Hayden et al., 2006; Jadad and McQuay, 1996; Jadad et al., 2000; Mallen et al., 2006; Moher et al., 2007; Whiting et al., 2005). This undermines the public's ability to be confident in the integrity of the process.

Reporting standards provide transparency by requiring extensive discussion on the methods used to conduct the review in sufficient detail to replicate the results. In 1999 and 2000, QUOROM (Quality of Reporting of Meta-analyses) and MOOSE (Meta-analysis Of Observational Studies in Epidemiology) reporting standards were published to improve the quality of meta-analyses, although neither set of standards has become widely adopted (Moher et al., 2007). CONSORT (Consolidated Standards of Reporting Trials) has simplified the task of summarizing evidence from randomized controlled trials (Moher et al., 1999; Stroup et al., 2000).

BUILDING THE PROGRAM'S FOUNDATION

This section considers how best to approach building the Program based on the foundational principles outlined above. The section begins with a brief review of programs in other countries and then examines three alternative models for the United States.

International Approaches to Identifying Effective Services

Many countries have developed programs to examine the effectiveness of clinical services. In Europe, 16 countries have at least one publicly affiliated agency responsible for assessing clinical effectiveness. Australia, Canada, and Singapore, among other countries, also have clinical effectiveness programs. As with the efforts made by various agencies and parties to assess clinical effectiveness in the United States, over the past three to four decades efforts elsewhere in the world have been prompted by concern with the high cost of medical interventions, as well as concern about the unsubstantiated benefits of widely disseminated clinical practices (Jonsson, 2002; Oliver et al., 2004).

The European Community (EC) has promoted priority setting, effectiveness assessments, and information sharing and the dissemination of results since 1994 (Velasco-Garrido and Busse, 2005). Health technology assessment has been a specific priority of the EC since 2004. The EC established the European Network for Health Technology Assessment (EUnetHTA) in 2006 to promote better coordination of national efforts (Kristensen and the EUnetHTA Partners, 2006). This Europe-wide initiative serves as an umbrella effort to make sure that there is no duplication of efforts and to bring up standards across individual countries and agencies.

Scope, Priority Setting, and Evidence Assessments in Selected National Programs

Systematic, detailed information on the operations of most national clinical effectiveness programs is limited, and studies assessing and comparing the impacts of these programs are even more limited (Oliver et al., 2004). The documentation and evaluation of national programs assessing clinical effectiveness that are available point to both the growth in capacity over time and the need for processes that are more consistent, transparent, and evidence based (Draborg and Gyrd-Hansen, 2005; García-Altés et al., 2004; Velasco-Garrido and Busse, 2005). The commit-

tee has not undertaken an in-depth study of international models for developing knowledge about clinical effectiveness and this brief overview does not endorse any country's particular approach.

The effectiveness review programs in Australia, Canada, Denmark, France, Germany, and the United Kingdom[2] assess a broad range of clinical services, including drugs, devices, tests, imaging procedures, preventive services, and surgical procedures (Table 6-1). The programs in Australia, Canada, Germany, and the United Kingdom assess both clinical effectiveness and cost-effectiveness (Table 6-2). In Australia, evidence of the comparative effectiveness of new drugs, devices, and procedures, including comparative cost-effectiveness, must be assessed before the national health insurance program will approve coverage. Manufacturers are required to submit extensive documentation on the effectiveness of their products to facilitate the assessment. In Canada, a national agency coordinates clinical and economic assessments and provides participating provincial and other public pharmaceutical benefits plans with coverage recommendations Canadian (CADTH, 2006). A governing board, composed of federal and regional health officials, selects which topics are to be assessed. In England and Wales, the National Institute for Health and Clinical Effectiveness (NICE), a special health authority within the National Health Service (NHS), assesses effectiveness. In Scotland, two organizations provide advice to the local health authorities within NHS Scotland: the Scottish Medicines Consortium, which reviews new drugs and new indications for the use of existing drugs for clinical effectiveness and cost-effectiveness, and the Scottish Intercollegiate Guidelines Network (SIGN), which develops and disseminates recommendations for effective clinical practices.

TABLE 6-1 Focus of Selected National Efforts to Identify Effective Health Services

Country	Drugs	Devices[a]	Preventive Services	Surgical Procedures[b]
United States	√		√	
Australia	√	√	√	√
Canada	√	√	√	√
Denmark	√	√	√	√
France	√	√	√	√
Germany	√	√	√	√
Scotland	√	√	√	√
England and Wales	√	√	√	√

[a] Includes diagnostic and therapeutic devices (e.g. ultrasound machines, stents, and inhaler devices)
[b] Includes the assessment of operating techniques, the use of surgical equipment for a specific procedure, and comparative effectiveness of surgical procedures.
SOURCE: Australian Safety & Efficacy Register of New Interventional Procedures – Surgical (2005); CADTH (2006); Canadian Task Force on Preventive Health Care (2005); Department on Health and Ageing (2006); Haute Autorité de santé (2007); Institute for Quality and Efficiency in Health Care (2007); National Board of Health (2007); National Health and Medical research Council (2006); NICE (2007); SIGN (2007).

[2] England and Wales have a separate program from Scotland.

TABLE 6-2 Key Features of National Clinical Effectiveness Programs in Australia, Canada, and England and Wales

National Organization (Country)	Scope of Review	Entities that Select Topics and Set Priorities	Entities That Perform Evidence Assessments	Types of Decisions
Pharmaceutical Benefits Advisory Committee (Australia)	Comparative clinical and cost-effectiveness of drugs	Manufacturers seeking coverage of new drugs submit application for review.	Internal and external organizations. Manufacturers and other third parties must submit detailed applications to support coverage review.	Coverage (advisory to Minister of Health and Ageing)
Medical Services Advisory Committee (MSAC) (Australia)	Safety, effectiveness, and cost effectiveness of new medical technologies and procedures	Medical profession, industry, or others seeking coverage for new medical technology or procedure submit application; MSAC prioritizes reviews.	External health technology assessment organizations advised by internal panels of MSAC members, experts, and consumers.	Coverage (advisory to Minister of Health and Ageing)
Canadian Agency for Drugs and Technologies in Health (Canada)	Clinical and cost effectiveness of drugs, devices for diagnosis and treatment, procedures, and health services management	Board of Directors (Deputy Health Ministers from federal, provincial, and territorial health agencies) selects topics.	Internal and external organizations; activities of seven provincial health technology assessment organizations are coordinated.	Coverage recommendations for drugs; advisory for other services
NICE (England and Wales)	Clinical and cost effectiveness of drugs, devices, diagnostics, surgical procedures, and health promotion interventions	Individuals and groups[a] may propose topics. Department of Health selects topics.	External groups perform initial assessment; expert committees are convened to do final assessment with internal staff support.[b]	Coverage, development of guidelines and clinical audit methods

[a] Includes health professionals, patients and the general public, clinical directors within the Department of Health, manufacturers, and the National Horizon Scanning Centre of the University of Birmingham (a group that tracks emerging technologies).
[b] Manufacturers may submit an initial assessment which is then reviewed and critiqued by an external review group.
SOURCE: Lopert (2006); Miller (2006); Sanders (2002).

Relevance to the United States

The countries listed in Table 6-1 differ from the United States in that they have government-sponsored health coverage. Yet, none of those national programs supports a health system that exceeds the scope of current U.S. federal expenditures on health—an estimated $645 billion in 2005—for Medicare, Medicaid, the State Children's Health Insurance Program, the U.S. Department of Defense, the Veterans Health Administration, and the Indian Health Service. Moreover, the United States spends more per capita on health care than any other country. In 2002, U.S. per capita health spending was $5,267; 53 percent more than any other country (Anderson et al., 2005).Thus, despite smaller expenditure bases, these national systems have chosen to make substantial investments to identify the most effective clinical services and apply such knowledge to promote and improve health outcomes. Many of them also take explicit account of the cost-effectiveness of particular clinical services to conserve and optimize their programs' finite finan-

cial resources. Notably, these national systems use relatively centralized coverage-oriented programs both to improve the investment of public resources in health care (e.g., the Pharmaceutical Benefits Advisory Committee in Australia) and to ensure the availability of effective new technologies throughout a national system (e.g., NICE in England and Wales).

It is difficult to generalize about the impact of national technology assessment programs on the adoption of new clinical interventions. One recent study that examined the rates of diffusion of new clinical technologies in 10 countries found mixed results for the adoption of particular technologies across countries. Still, the presence of a clinical effectiveness report or some other form of guidance was consistently associated with the increased diffusion of the technology (as was above-average per capita spending on health care) (Packer et al., 2006).

Another insight from the international experience with programs that assess clinical effectiveness is that the mere development and publication of information, even by the most authoritative sources, are not in and of themselves sufficient to ensure changes in policy and practice (Battista, 2006; Oliver et al., 2004). National programs have moved in the direction of increasing the transparency of their assessment processes, placing a greater emphasis on the dissemination and communication of the results of assessments, and in some cases encouraging greater consumer involvement. In structuring a program uniquely suited to U.S. circumstances, the United States can learn from the history of and progress that other countries have made.

Alternative Models for a U.S. National Clinical Effectiveness Assessment Program

The committee considered three approaches to establishing the Program infrastructure: maintaining the status quo and two alternatives (described below). Table 6-3 compares key aspects of the status quo with the two proposed alternatives: a central agency and a hybrid model. Both alternatives to the status quo would require that the Program substantially scale up resources, develop rigorous methodological and reporting standards (including common terminology), and institute protections against bias due to conflict of interest.

Status Quo

As the previous chapters described, the committee found convincing evidence that systematic reviews and clinical guidelines are often of poor quality, lacking scientific rigor and objectivity. Under the status quo, systematic reviews and clinical guidelines are produced by numerous public and private organizations with little or no coordination, minimal quality controls, inconsistent terminology, inadequate transparency, and without concerted attention to the priorities of all types of consumers, patients, and other stakeholders. Perhaps as a consequence, while many important topics remained unexamined, there is unnecessary duplication of effort in assessments of new and emerging technologies. No one agency or organization in the United States evaluates from a broad, national perspective the effectiveness of new as well as established health interventions for all populations, children as well as elderly people, women as well as men, and ethnic and racial minorities.

Central Agency Model

The first alternative to the status quo, coined the “central agency model,” is a single, highly centralized entity, such as an executive branch agency or a division of an executive agency. It would have broad authority to fund, carry out, and control the full range of analytic tasks: setting

priorities for systematic reviews, producing systematic reviews, and developing clinical guidelines—all in accordance with mandatory Program standards. Some or all of the Program's procedures could be based in statute (e.g., mandatory priority setting criteria). The agency would be led by executive level staff who would oversee Program activities with support from an extensive Program staff.

Hybrid Model

The second alternative to the status quo, referred as the "hybrid" model, builds on current private and public sector capacity but gives the Program the authority and sufficient funding to develop process and reporting standards for, to set priorities for, and to sponsor standards-based systematic reviews of high priority topics. The Program's role regarding clinical guideline development would be threefold: (1) developing (or endorsing) rigorous but voluntary guidelines standards, (2) promoting voluntary compliance with guideline standards, and (3) providing a forum for resolving conflicts between existing guidelines. An independent advisory board would oversee the Program. A group of core staff would be needed but the Program would rely extensively on outside experts and organizations.

TABLE 6-3 Alternative Approaches to Organizing the Program: Administrative Structure and Primary Functions

Organizational Feature or Function	Status Quo	Agency Model	Hybrid Approach
Structure			
Administrative infrastructure	No change	Infrastructure is sufficient to support significant expansion in evidence assessment and to develop standards for evidence assessments, clinical guidelines, and bias protections. Executive staff oversee the Program.	Infrastructure is sufficient to support significant expansion in and to develop standards for systematic reviews, clinical guidelines, and bias protections. An independent advisory board oversees the Program. Membership of the board includes diverse public and private sector expertise.
Degree of program control over clinical effectiveness assessment process	There is no change, except when sponsored by the AHRQ Effective Health Care Program.	High. Mandatory standards and processes. In-house staff oversee and conduct key functions for priority setting, evidence reviews, and clinical recommendation development.	Mixed. Control over priority setting and to a large extent over systematic review functions, which must meet standards and bias protections. No direct control over clinical recommendation development, though standards set.
Primary functions			
Setting research priorities	Multiple public and private entities set program- or mission-specific priorities. AHRQ sets priorities as directed by the secretary of the U.S. Department of Health and Human Services.	Agency establishes priorities for systematic reviews of clinical effectiveness and clinical guidelines. Process is based in statute and provides for public and stakeholder input.	Priority Setting Advisory Committee (PSAC) establishes priorities for systematic reviews of clinical effectiveness (with public and stakeholder input). PSAC includes a broad mix of expertise and interests to minimize bias due to conflicts of interest.
Assessing evidence	Multiple, independent organizations operating without oversight. No standardized mechanisms for quality assurance and quality control.	Conducted by in-house staff and outside organizations in accordance with program standards. Stronger protections against bias.	Conducted in accordance with program standards. Stronger protections against bias.
Developing clinical guidelines/ recommendations	Multiple, independent organizations operating without oversight. Multiple, voluntary practice guidelines are available. No standardized mechanisms for quality assurance and quality control; claims of evidence base not necessarily supported by methods.	Developed by in-house staff and outside organizations in accordance with program standards. Stronger protections against bias.	Multiple, independent organizations operating without oversight. Program promotes use of voluntary standards. No direct protections against bias in voluntary activities.

Comparing the Agency and Hybrid Models

Table 6-4 compares the committee's assumptions about the alternative models' likely adherence to the guiding principles outlined earlier in Box 6-1. From a hypothetical perspective, a highly centralized effort (i.e., the agency model) appears to be more likely to offer maximum control over both evidence assessment and guideline development and, thus, theoretically a greater likelihood of optimizing the key principles. This model, however, is also likely to be the most costly, to generate more political opposition, and also to take more to time to establish than an approach that builds on current capacity. With the burgeoning array of new devices, medical technologies, and biological therapies, time is of the essence.

The critical difference between the hybrid Program infrastructure and the central agency model, are the entities that would formulate clinical guidelines. In both models, the quality of systematic reviews could be addressed through the application of rigorous process and reporting standards. The standards could be newly created or already developed standards that are endorsed by the Program. In the central agency model, the Program itself would oversee clinical guideline development as well as the systematic reviews. Under the hybrid approach, the Program would sponsor standards-based systematic reviews of high priority topics by outside experts. In contrast with the agency model, the hybrid model assumes that existing independent entities—professional medical societies, payers, practice measurement groups, and others—would continue to develop clinical guidelines. The Program would actively encourage these organizations to voluntarily adopt Program standards for guideline development.

The agency and hybrid alternatives also differ with respect to the administrative infrastructure required to support the Program. Under the agency model, an extensive in-house staff would support or carry out key functions including priority setting, evidence reviews, and clinical guideline development. The hybrid approach would require fewer staff and build on current, outside capacity. The hybrid model also calls for an independent Priority Setting Advisory Committee, as described in Chapter 3, to establish and regularly update Program priorities for systematic review.

TABLE 6-4 Summary Assessment of Organizational Alternatives Based on Committee Principles

Principles	Status Quo	Agency Model	Hybrid Model
	No change	*Centralizes responsibility in an expanded or new agency which determines priorities, funds, produces, and sets mandatory standards and language for both systematic reviews and clinical recommendations/ guidelines. Responsible for making clinical guidelines and recommendations.*	*A national Program determines priorities (with public input), funds, and sets mandatory standards and language for systematic reviews. External groups and individuals produce systematic reviews. Establishes voluntary standards for clinical recommendations/guidelines. Existing organizations produce clinical guidelines and recommendations.*
Accountability—*Parties are directly responsible for meeting and demonstrating compliance with minimum standards*	*Poor*. Systematic reviews and guidelines are often proprietary or available only to members. When publicly available, methods used often lack complete documentation.	*Moderate to high*. Central agency is directly responsible for and reports on compliance. Congress provides oversight.	*Moderate to high*. Program is directly responsible for priority setting and systematic reviews. Reliance on disclosure of compliance with common standards and end user preference for guidelines produced according to standards.
Consistency—*Standardized and predictable methods*	*Poor*. Systematic reviews and clinical recommendations may not use standardized, evidence-based methods.	*High*. Standardization of methods is accomplished with a unified management structure.	*Moderate to high*. Funding mechanism for systematic reviews requires standardization of methods. Reliance on disclosure of compliance with common standards and end user preference for guidelines produced according to standards.
Efficiency—*Avoids waste and unnecessary duplication*	*Poor*. Redundant and conflicting evidence reviews and guidelines are common.	*Moderate to high.* Depends on effective and well-targeted implementation.	*Moderate to high.* Unnecessary duplication of priority setting and systematic reviews is reduced. Potential for duplication of clinical recommendations remains.
Feasibility—*Capable of operating in the real world*	*High.* No change from current practice. But without additional funding, output will be relatively low or unpredictable from year to year.	*Poor.* Political support seems unlikely given high cost, new bureaucracy, and assumption of some responsibilities previously in the private sector (i.e., making clinical recommendations). Private sector organizations may strongly resist the agency's takeover of some of their current activities. Will require larger professional-technical workforce.	*Moderate.* Requires new or expanded infrastructure and increased expenditures. May face political resistance among some affected stakeholders. Will require larger professional-technical workforce but more will be accomplished.

TABLE 6-4 Summary Assessment of Organizational Alternatives Based on Committee Principles *cont'd*

Principles	Status Quo	Agency Model	Hybrid Model
Objectivity—*Evidence-based and without bias; conflict of interest is minimized*	*Poor*. Voluntary and conflicting standards, inconsistently applied.	*High*. Integrated process and autonomous operational structure supports enforcement of standards.	*Moderate to high*. Program products must meet common standards for conflict of interest, priority setting, and production of systematic reviews that minimize statistical bias. Reliance on disclosure of compliance with common standards and end user preference for guidelines produced according to standards.
Responsiveness—*Addresses information needs of decision makers (i.e., consumers, health professionals, payers and purchasers, etc.)*	*Poor*. No national priorities. Existing reviews do not address many patient populations (e.g., children, elderly) or the full continuum of services. Information on the comparative effectiveness of health services is largely lacking.	*Moderate to high*. Significant start up time required. Decision makers might have input into priority setting. Ability to respond depends on government oversight.	*High*. Actively seeks input from decision makers regarding priority topics for systematic reviews. Fewer procedural requirements/steps shorten response time.
Scientific rigor—*Methods minimize bias, are reliable, and completely reported*	*Poor*. Evidence-based methods may not be used; errors and poor documentation are common.	*Moderate to high*. Required by Program standards; program funding ensures that resources are available to support rigorous work. But performance will depend on well-trained staff with requisite scientific skills.	*Moderate to high*. Process maximizes likelihood that priority setting and systematic reviews would meet scientific standards. Reliance on disclosure of compliance with common standards and end user preference for guidelines produced according to standards.
Transparency—*Methods explicitly defined, consistently applied, and publicly available*	*Poor*. Appropriate documentation is often lacking. Information is often proprietary or not publicly available.	*High*. Required by Program standards and subject to federal disclosure requirements.	*Moderate* to high. Standards are publicly available. Reliance on disclosure of compliance with common standards and end user preference for guidelines produced according to standards.

RECOMMENDATIONS FOR BUILDING THE PROGRAM INFRASTRUCTURE

This report has outlined an urgent imperative for immediate action to change how the nation marshals clinical evidence and applies it to identify the most effective clinical interventions. The nation's annual multibillion dollar investment in biomedical research and innovation has provided many important insights into human health and disease, yet only a fraction of one percent of U.S. spending on biomedical research is invested in identifying what constitutes sound and reliable evidence of the most effective health services (Emanuel et al., 2007). Evidence assessment (i.e., systematic review) is central to scientific inquiry into what is known and not known about what works in health care. The previous chapters outlined the committee's rationale and recommendations for three essential Program functions: priority setting, evidence assessment (systematic review), and developing standards for clinical guidelines. The following presents the committee's recommendations for establishing an infrastructure for organizing the three functions. The committee's complete set of recommendations are summarized in Box 6-2.

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.**

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 urges that the Program incorporate substantial stakeholder involvement, develop (or endorse) methodologic and reporting standards for systematic reviews and clinical guidelines, and adopt rigorous standards for minimizing bias and conflict of interest in the Program.

An Independent Forum

Under the status quo, there are many conflicting clinical practice guidelines. Consumers, patients, health professionals, and others struggle to learn which guideline is appropriate for which circumstances. The committee suggests that the Program sponsor ongoing, public meetings that are organized to help resolve differences between conflicting clinical guidelines and recommendations. Such an independent forum would provide an important public service.

Program Evaluation

The Program must be accountable to Congress and the public. The committee recommends that the Clinical Effectiveness Advisory Board routinely evaluate the Program to ensure that it is fulfilling its purpose effectively and also submit an annual report on its activities and accomplishments to Congress.

UNANSWERED QUESTIONS

As Chapter 1 described, the scope of this study did not address several critical concerns that merit attention: where to place the Program and whether it should be public, private, or a public-private collaboration; program costs and sources of program funding; technical methods including the use of cost data and cost-effectiveness methods in assessing effectiveness; knowledge transfer and how to assure adherence to guidelines; how to reflect patient values and preferences in clinical guidelines, and legal issues.

BOX 6-2 Committee 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.

REFERENCES

AGREE Collaboration. 2001. *Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument* www.agreecollaboration.org (accessed December 8, 2006).

AHRQ (Agency for Healthcare Research and Quality). 2007. *Effective Health Care—Home Page* http://effectivehealthcare.ahrq.gov/ (accessed August 7, 2007).

America's Health Insurance Plans. 2007. *Setting a higher bar: We believe there is more the nation can do to improve quality and safety in health care.* Washington, DC: AHIP.

Anderson, G. F., P. S. Hussey, B. K. Frogner, and H. R. Waters. 2005. Health spending in the United States and the rest of the industrialized world. *Health Affairs* 24(4):903-914.

Australian Safety & Efficacy Register of New Interventional Procedures – Surgical. 2005. *Annual Report.* Melbourne, Australia: Royal Australasian College of Surgeons.

Battista, R. N. 2006. Expanding the scientific basis of health technology assessment: A research agenda for the next decade. *International Journal of Technology Assessment in Health Care* 22(3):275-280

Bhandari, M., F. Morrow, A. V. Kulkarni, and P. Tornetta. 2001. Meta-analyses in orthopaedic surgery: A systematic review of their methodologies. *Journal of Bone and Joint Surgery* 83A:15-24.

BCBSA (Blue Cross and Blue Shield Association). 2007a. *Blue Cross and Blue Shield Association's Technology Evaluation Center* http://www.bcbs.com/tec/index.html (accessed January 18, 2007).

———. 2007b. Blue Cross and Blue Shield Association proposes payer-funded institute to evaluate what medical treatments work best. http://www.bcbs.com/news/bcbsa/blue-cross-and-blue-shield-association-proposes-payer-funded-institute.html (accessed May 2007).

CADTH (Canadian Agency for Drugs and Technologies in Health). 2006. *Health technology assessment* http://www.cadth.ca/index.php/en/hta (accessed March 28, 2007).

Canadian Task Force on Preventive Health Care. 2005. *Evidence-Based Clinical Prevention* http://www.ctfphc.org/ (accessed March 28, 2007).

Cochrane Collaboration. 2007. *The Cochrane Collaboration: The reliable source of evidence in health care* http://www.cochrane.org/ (accessed January 18, 2007).

Congressional Budget Office. 2007. Research on the comparative effectiveness of medical treatments: Options for an expanded federal role. *Testimony by Director Peter R. Orszag before House Ways and Means Subcommittee on Health*, http://www.cbo.gov/ftpdocs/82xx/doc8209/Comparative_Testimony.pdf (accessed June 12, 2007).

Delaney, A., S. M. Bagshaw, A. Ferland, B. Manns, and K. B. Laupland. 2005. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 9:R575–R582.

Department on Health and Ageing. 2006. *About us: Our role* http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-overview.htm (accessed March 28, 2007).

Draborg, E., and D. Gyrd-Hansen. 2005. Time-trends in health technology assessments: An analysis of developments in composition of international health technology assessments from 1989 to 2002 *International Journal of Technology Assessment in Health Care* 21(4):492-498.

ECRI. 2006. *About ECRI* http://www.ecri.org/About_ECRI/About_ECRI.aspx (accessed January 31, 2007).

Emanuel, E. J., V. R. Fuchs, and A. M. Garber. 2007. Essential elements of a technology and outcomes assessment initiative. *JAMA* 298(11):1323-1325.

García-Altés, A., S. Ondategui-Parra, and P. J. Neumann. 2004. Cross-national comparison of technology assessment processes. *International Journal of Technology Assessment in Health Care* 20(3):300-310

Glenny, A. M., M. Esposito, P. Coulthard, and H. V. Worthington. 2003. The assessment of systematic reviews in dentistry. *European Journal of Oral Sciences* 111:85-92.

Gøtzsche, P. C., A. Hrobjartsson, K. Maric, and B. Tendal. 2007. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 298(4):430-437.

GRADE Working Group. 2004. Grading quality of evidence and strength of recommendations. *BMJ* 328(7454):1490-.

Gray, B. H. 1992. The legislative battle over health services research. *Health Affairs* 11(4):38-66.

Gray, B. H., M. K. Gusmano, and S. Collins. 2003. AHCPR and the changing politics of health services research. *Health Affairs*:w3.283.

Hayden, J. A., P. Cote, and C. Bombardier. 2006. Evaluation of the quality of prognosis studies in systematic reviews *Annals of Internal Medicine* 144:427-437.

Haute Autorité de santé. 2007. *About HAS* http://www.has-sante.fr/portail/display.jsp?id=c_5443&pcid=c_5443 (accessed March 28, 2007).

Hayes, W. S. 2006. *Healthcare policy: Applying the evidence (PowerPoint presentation to the HECS Committee meeting, November 7, 2006)*. Washington, DC.

Higgins, J. T., and S. Green. 2006. *Cochrane handbook for systematic reviews of interventions 4.2.6 [updated September 2006], The Cochrane Library, Issue 4, 2006.* Chichester, UK: John Wiley & Sons, Ltd.

Institute for Quality and Efficiency in Health Care. 2007. *About us* http://www.iqwig.de/about-us.21.en.html (accessed March 28, 2007).

IOM (Institute of Medicine). 1992. *Setting priorities for health technologies assessment: A model process*. Edited by Donaldson, M. S. and H. C. Sox. Washington, DC: National Academy Press.

———. 1995. *Setting priorities for clinical practice guidelines*. Edited by Field, M. J. Washington, DC: National Academy Press.

———. 2001. Crossing the quality chasm: A new health System for the 21st Century. Washington, DC: National Academies Press.

———. 2007. *Learning what works best: The nation's need for evidence on comparative effectiveness in health care*. Washington, DC: The National Academies Press. http://www.iom.edu/ebm-effectiveness.

Jadad, A. R., and H. J. McQuay. 1996. Meta-analyses to evaluate analgesic interventions: A systematic qualitative review of their methodology. *Journal of Clinical Epidemiology* 49:235-243.

Jadad, A. R., M. Moher, G. P. Browman, L. Booker, C. Sigouin, M. Fuentes, and R. Stevens. 2000. Systematic reviews and meta-analyses on treatment of asthma: Critical evaluation. *BMJ* 320:537-540.

Jonsson, E. 2002. Development of health technology assessment in Europe. *International Journal of Technology Assessment in Health Care* 18(2):171-183.

Kristensen, F. B., and the EUnetHTA Partners. 2006. EUnetHTA and health policy-making in Europe. *Eurohealth* 12(1).

Kupersmith, J., N. Sung, M. Genel, and et al. 2005. Creating a new structure for research on health care effectiveness. *J of investigative medicine* 53(2):67-72.

Lexchin, J., L. A. Bero, B. Djulbegovic, and O. Clark. 2003. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *BMJ* 326:1167–1170.

Lopert, R. 2006 (unpublished). *Pharmacoeconomics and drug subsidy in Australia--an account of the Australian approach*.

Mallen, C., G. Peat, and P. Croft. 2006. Quality assessment of observational studies is not commonplace in systematic reviews. *Journal of Clinical Epidemiology* 59:765–769.

Medicare Payment Advisory Commission. 2007. Chapter 2: Producing comparative effectiveness information.29-54, http://www.medpac.gov/documents/Jun07_EntireReport.pdf (accessed June 2007).

Miller, W. 2006. *Value-based coverage policy in the US and UK: Different paths to a common goal.* Washington, DC: National Health Policy Forum.

Moher, D., D. J. Cook, S. Eastwood, I. Olkin, D. Rennie, D. F. Stroup, and the QUOROM Group. 1999. Improving the quality of reports of meta-analyses of randomized controlled trials: The QUOROM statement. *Lancet* 354:1896-1900.

Moher, D., J. Tetzlaff, A. C. Tricco, M. Sampson, and D. G. Altman. 2007. Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine* 4(3):447-455.

National Board of Health. 2007. *Danish Centre for Health Technology Assessment* http://www.sst.dk/Planlaegning_og_behandling/Medicinsk_teknologivurdering.aspx?lang=en (accessed March 28, 2007).

National Health and Medical Research Council. 2006. *Role of the NHMRC* http://www.nhmrc.gov.au/about/role/index.htm (accessed March 28, 2007).

National Research Council. 2004. *Eliminating health disparities: Measurement and data needs*. Edited by Ver Ploeg, M. and E. Perrin. Washington, DC: The National Academies Press.

NICE (National Institute for Health and Clinical Excellence). 2007. *About technology appraisals* http://www.nice.org.uk/page.aspx?o=202425 (accessed March 28, 2007).Oliver, A., E. Mossialos, and R. Robinson. 2004. Health technology assessment and its influence on health-care priority setting *International Journal of Technology Assessment in Health Care* 20(1):1-10

NQF (National Quality Forum). 2006. *Organizational values* http://qualityforum.org/about/values.asp (accessed August 6, 2007).

Packer, C., S. Simpson, and A. Stevens (on behalf of EuroScan: the European Information Network on New and Changing Health Technologies). 2006. International diffusion of new health technologies: A ten-country analysis of six health technologies *International Journal of Technology Assessment in Health Care* 22(4):419-428.

Sanders, J. M. 2002 Challenges, Choices and Canada. *International Journal of Technology Assessment in Health Care* 18(2):199-202.

Shojania, K. G., M. Sampson, M. T. Ansari, J. Ji, S. Doucette, and D. Moher. 2007. How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine* 147:224-233.

Shortell, S. M., T. G. Rundall, and J. Hsu. 2007. Improving patient care by linking evidence-based medicine and evidence-based management. *JAMA* 298(6):673-676.

SIGN (Scottish Intercollegiate Guidelines Network). 2007. *Guideline Development Programme* http://www.sign.ac.uk/guidelines/development/index.html (accessed March 28, 2007).

Simpson, L. 2004. Lost in translation? Reflections on the role of research in improving health care for children. *Health Affairs* 23(5):125-130.

Steinberg, E. P., and B. R. Luce. 2005. Evidence based? Caveat emptor! *Health Affairs* 24(1):80-92.

Stroup, D. F., J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, D. Rennie, D. Moher, B. J. Becker, T. A. Sipe, and S. B. Thacker for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. 2000. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA* 283(15):2008-2012.

Velasco-Garrido, M., and R. Busse. 2005. *Health technology assessment: An introduction to objectives, role of evidence, and structure in Europe.* Brussels, Belgium: WHO European Observatory on Health Systems and Policies.

Whiting, P., A. W. Rutjes, J. Dinnes, J. B. Reitsma, P. M. Bossuyt, and J. Kleijnen. 2005. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *Journal of Clinical Epidemiology* 58:1-12.

Wilensky, G. R. 2006. Developing a center for comparative effectiveness information. *Health Affairs*: w572.