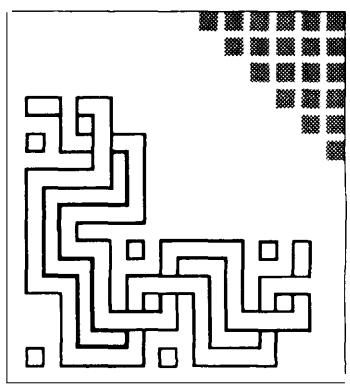
Summary 1

1

reproving the health of Americans through the discovery and implementation of new medical technologies [1] has been an explicit goal of the federal government for over a century. Since the 1970s, however, the government has also underwritten a less visible effort-the attempt to identify which health care interventions, among those in current use, work best.

The justification for most medical practices used in the United States today rests on the experience and expertise of clinicians and patients rather than on objective evidence that these practices can measurably improve people's health. Compiling objective evidence is considered by many people to be costly and unnecessary. It is also highly controversial, because the evidence might be applied in ways that would limit individuals' choices of medical treatments.

But the reliance on personal experiences as the basis of existing medical practices has been increasingly questioned. Evidence has been slowly accumulating that suggests that even well-accepted and very common technologies, such as routine chest x-rays, can be ineffective, that a substantial number of medical and surgical procedures are performed for inappropriate reasons, and that different regions supply very different amounts of medical care, with



[1] The congressional Office of Technology Assessment defines "medical technology" as comprising drugs, devices, procedures, and the organizational and support systems within which medical care is delivered (780). Most of this report discusses examples and issues from the medical technology arena. However, the issues are also applicable to health care interventions more broadly—i.e., not only specific technologies and sets of [technologies from clinical care, but also interventions as diverse as lead abatement programs and efforts to implement clinical practice guidelines. "Medical technologies" and "health care interventions" are thus sometimes used interchangeably in this context.

very different costs, despite apparently similar levels of underlying need. At the same time, the American health care system is frequently criticized for being the costliest in the world, despite the fact that the United States lags behind many other nations in basic measures of population health, such as life expectancy and infant mortality.

The basic rationale for the current federal effort to identify which existing health care technologies work best has been the hope that the results of this effort can increase not only the benefits of health care but also the value. As a number of advocates have argued, if a particular use of a technology is ineffective or unnecessary, eliminating that use should benefit patients and payers alike.

Many of the proposals for reforming the health care system currently being debated by federal and state legislatures rely on research into medical effectiveness and cost-effectiveness, along with clinical practice guidelines backed by this research, to support the changes they envision. These proposals include strategies such as:

- linking insurance benefits to the effectiveness and cost-effectiveness of particular technologies and services;
- changing the legal standard of care to permit physicians to be protected from malpractice suits if they have followed clinical practice guidelines;
- increasing the use of managed care (which implies the greater use of guidelines on which to base internal management strategies); and
- using "report cards" to judge and compare health care providers and plans, a strategy that uses published indicators intended to represent how well those providers adhere to effective care practices.

These strategies rest on the expectation that research will identify which health care technologies work best.

The focal point of the federal government's medical effectiveness research effort is the Agency for Health Care Policy and Research (AHCPR), in the U.S. Department of Health and Human Services (DHHS). Congress created AHCPR in 1989 specifically to further the evaluation of existing clinical practice.[2] AHCPR was charged with conducting research to identify effective care, developing guidelines for clinical practice based on this research, and disseminating knowledge about effective care patterns (Public Law 101-239). When AHCPR was reauthorized by Congress in 1992, its mandate was changed slightly to reflect the heightened congressional interest in identifying cost-effective, as well as simply effective, care. AHCPR's mandate now also requires the agency to consider the costs of different care patterns considered in clinical practice guidelines and to include cost-effectiveness analyses in its assessments of individual technologies (Public Law 102-410).

The potential of AHCPR's research and clinical guidelines activities to help solve some of the problems of the health care system, along with the increasing federal investment in those and related activities, led Congress in 1992 also to request this Office of Technology Assessment (OTA) study of the effort and its ability to realize its potential (box 1-A).

Although AHCPR plays a special role in evaluating the worth of health technologies in current use, it is by no means the only federal agency engaged in relevant activities. The focus of this report is on the spectrum of federal activities that address three components of the evaluation of health care technologies:

1. research into the effectiveness of health care technologies in current use,
2. analysis of the comparative cost-effectiveness of alternative technologies, and
3. the broader assessment of existing health care technologies for policy purposes,

[2] AHCPR absorbed the National Center for Health Services Research, which had sponsored much of the general health services research in the 1970s and 1980s that ultimately led to the medical effectiveness initiative.

BOX 1-1: The Origins of the OTA Study

The Office of Technology Assessment (OTA), an analytic support agency of the U S Congress, undertakes studies at the request of committees of Congress OTA published several reports on the conduct of clinical research and health technology assessment in the late 1970s and early 1980s (778,779,780,783,784) Since 1983, however, most OTA health-related reports have been assessments of specific technologies, and technology-related health care issues, rather than studies of the process and methods of health technology assessment

New approaches to evaluating the effectiveness and cost-effectiveness of health technologies, and congressional discussion surrounding the contemplated reauthorization of the Agency for Health Care Policy and Research in 1992, prompted Congress to ask OTA to revisit the issues of health technology assessment and research In July 1992, Senators Kennedy and Hatch, on behalf of the Labor and Human Resources Committee, asked that OTA 'conduct an evaluation of the field of health technology assessment, Identify strengths and weaknesses of current efforts, and outline options which may help focus future efforts and resources" (427) Types of activities to be covered in this evaluation were "literature synthesis, outcomes research, cost-effectiveness analysis, practice guideines development, and others."'

Senator Grassley, of the congressional Technology Assessment Board, and Congressman Dlngell, on behalf of the House Committee on Energy and Commerce, also sent letters supporting an OTA study of this topic They echoed the concerns expressed by Senators Kennedy and Hatch and emphasized the Importance of being able to develop "accurate Information on the value of various procedures and medical technologies" so that "payers, prowders and consumers can make efficient decisions regarding care" (1 76,294)

The initiation of the OTA study was approved by OTA's congressional Technology Assessment Board in August 1992 The study began on October 1 of that year

SOURCE Off Ice of Technology Assessment 1994

The report is especially concerned with clinical practice guidelines. In the context of public policy, clinical practice guidelines can be viewed as a unique form of health technology assessment that is intended to affect clinical decisions directly. as well as indirectly, through insurance payment or other policies that are linked to those guidelines. The primary goals of this report are:

- to assess the current state of the federal activities in these areas,
- to identify what can realistically be expected from investing in these activities, and
- to identify areas in which current efforts are especially weak or are missing important opportunities.

SUMMARY OF FINDINGS AND CONCLUSIONS

Health care can be improved at man y different levels. At the local level, physicians and other providers may attempt to improve the quality of the care they provide by altering their processes of care to enhance patient satisfaction, to adhere more closely to existing standards of effective care, and to improve the health of their patients.

Additional improvements in health care can be made at the level of the health care system overall. As the system improves its knowledge of which technologies and services work better than others, for which patients, and under which cir-

Medical technology is intrinsic to American health care, but most technologies currently in use have never been rigorously tested for their effectiveness or cost-effectiveness

cumstances, providers can use this knowledge to improve the care they give. Identifying "what works best" in health care at the policy level has four overlapping components:

1. The efficacy and safety of a health care intervention: whether a given intervention can, at least under ideal circumstances, improve some people's health.
2. The effectiveness of an intervention: whether including it in the repertoire of health care improves people's health under ordinary circumstances, in ordinary settings, and whether it generally improves health more than alternative interventions (comparative effectiveness).
3. The cost-effectiveness of an intervention: whether, compared with other alternatives, its combined economic and medical value makes it worth doing.
4. The overall impact of an intervention as it relates to the decisions that policy makers must make-i. e., health technology assessment. The policy decisions addressed by the technology assessment may be clinical policies, purchasing or payment policies, or public policies; depending on the needs of the policy makers, they may be restricted to concerns about effectiveness and cost-effectiveness or raise issues such as legal concerns, distributional effects, and effects on access to care.

Effectiveness Research

In the framework of this report, "effectiveness research" encompasses research efforts aimed at identifying broadly effective care, and efforts to develop and refine methods to support the identification of effective care.[4]

The federal government's medical effectiveness initiative, as reflected in the statutory charge to AHCPR and the agency's implementation of that charge, has emphasized some aspects of effectiveness research and de-emphasized others. The outstanding characteristics of the federal endeavor have been:

1. The federal effort has focused primarily on evaluating technologies and medical practices currently in use, rather than on the evaluation of new interventions.
2. It has emphasized the need for research that will permit generalizations about effectiveness to be made to populations and settings-elderly people, women, minorities, persons with disabilities or multiple health problems, and treatment settings such as health facilities not affiliated with teaching institutions—that have often been underrepresented in past efficacy studies.
3. It has stressed the use of outcome measures that assess factors that affect patients directly (e.g., physical and social functioning and pain), rather than intermediate clinical measures (e.g., laboratory test scores).

[3] "Efficacy" and "effectiveness" are useful to distinguish conceptually, but in practice they are closely related. For example, it is possible for a study to demonstrate both efficacy and effectiveness simultaneously if the population and settings included in the study are sufficiently diverse.

[4] "Outcomes research" is a popular phrase often used to describe this area of research, but because that phrase is also used to describe many other disparate activities as well, it has become a term laden with confusion and is rarely used in this report.

4. It has included the substantial use of tools other than prospective, randomized controlled trials (RCTs), emphasizing in particular the analysis of large administrative databases. It has not absolutely excluded the use of randomized and other controlled clinical studies, but much of the impetus for the field came from the expectation that for existing medical technologies, nonclinical research methods were often faster, cheaper, and more efficient.

One assumption underlying this effectiveness research effort has been that if the least effective practices can be identified and described through clinical practice guidelines, and the guidelines disseminated to clinicians and patients, it might be possible to raise the quality of health care while constraining or even reducing its costs. Early effectiveness research prompted great optimism about the possibilities of this research for identify - ing ineffective and inappropriate health care **prac-** tices. One line of research demonstrated the high proportions of inappropriate care that are sometimes provided, while another line of research demonstrated the great variations in clinical practice that occur. Together, they suggest that there is considerable room for improvement in health care that can be achieved by focusing on existing technologies and practices.

The goal of reducing inapproprlate and ineffective care is a worthwhile *and achievable one If implemented success fully, effectiveness research can improve the health care that patlents receive, but by ifself it cannot be expected to lower healfh care costs substantially*

Achieving these improvements, however, will not be as simple as is sometimes hoped, for three reasons. First, and perhaps most importantly, documenting variations in clinical practice does not itself provide information about which practices are the most effective. Producing this information requires additional directed, comparative research.

Second, reducing inappropriate care is not synonymous with reducing the costs of care. Many of the cited estimates of the amount of health care that is inappropriately provided and could be eliminated without affecting the quality of care in any way (e.g., 25 percent) are probably too high. Also. not all inappropriate care is a result of too much care. In some areas, it may be the low rates of a particular procedure that are inappropriate.

Third, the source of variations in clinical practice is not necessarily merely individual provider uncertainty about a technology's effectiveness, which could be abolished by simply presenting practitioners with good information or guidelines. Rather, physicians may often hold strong but opposing individual opinions, with some being enthusiasts for a procedure while others are more cautious users. Changing practice thus will require not merely better information but sufficient evidence, portrayed in a convincing way, to change opinions and actions.

Thus, while successfully implementing the findings of valid effectiveness research will probably improve the quality of health care, it will not necessarily reduce health care costs significantly. In fact, research on the effectiveness

of existing technologies and practices should be considered a good "buy" if it can succeed in improving health care while paying for its own research-related costs through targeted health system cost reductions.

As noted above, the focal point of federal effectiveness research is AHCPR, which was created in part specifically for this purpose. The stars of AHCPR's effectiveness research program are its Patient Outcomes Research Teams (PORTs).[5] These interdisciplinary research teams study specific medical conditions and the effectiveness of medical practices to diagnose, treat, and manage these conditions.

The PORTS, and other effectiveness research efforts supported by AHCPR, have made a number of contributions. Among the most important are:

- Raising the level of discussion about what is known, and what is not, about the effectiveness of treating particular diseases. PORT findings especially have helped clinicians and policy makers confront the inconsistencies in current medical practice, and they have created a fertile environment for new research on existing medical technologies and services.
- Developing and refining measures of health outcome that use patient self-assessments about health improvements, which have greatly aided researchers' ability to focus on the evaluation of outcomes of health interventions that most matter to patients. Effectiveness research has encouraged basic research on these tools, and it has contributed to an improved set of measures for assessing the outcomes of therapies for problems such as prostate disease, cataracts, and knee conditions.
- Highlighting the differences among medical practices shown to be effective and their use in particular populations of patients.
- Exploring new, potentially useful research applications of large pre-existing databases. Such applications include identifying potential participants for prospective studies; identifying rare adverse events; and combining clinical with administrative data, which offers possibilities for much richer descriptive information on the experiences of patients who have particular conditions and are undergoing particular treatments.
- Refining meta-analysis[6] and other systematic reviews of the literature and applying them more widely. Systematic reviews can reduce unnecessary and duplicative research, enable important information already available to gain broader exposure, clarify questions that need to be addressed with primary research, and reduce inconsistencies among literature reviews. PORT experience also shows, however, that if conducted inefficiently or without focus they can be costly and yield little.

While PORTS, and the federal effectiveness initiative more generally, have made contributions, their success has been qualified. Contrary to the expectations expressed in the legislation establishing AHCPR and the mandates of the PORTS, administrative databases generally have not proved useful in answering questions about the comparative effectiveness of alternative medical treatments. Administrative databases are very useful for descriptive purposes (e.g., exploring variations in treatment patterns), but the practical and theoretical limitations of this research technique usually prevent it from being able to provide credible answers regarding which technologies, among alternatives, work best.

Prospective comparative studies, and particularly RCTs, have been underused in the federal effectiveness initiative. The inability to follow up the questions highlighted by descriptive

[5] There were 14 active PORTs as of mid-1994. Four of those 14 PORTS end in the fall of 1994. An additional six new PORTs are starting up as the first four expire, under the revised "PORT-II" program.

[6] A meta-analysis is a systematic review of the results of previous clinical studies that includes a quantitative reanalysis of those studies' results.

medical effectiveness research with comparative clinical trials is one of the signal failures of the federal effectiveness effort.

Recently, AHCPR has made some changes in its research program, placing relatively more emphasis on primary data gathering and prospective studies in its effectiveness research agenda. These studies are not necessarily RCTs. though, and it is not yet clear whether the PORTS funded under the new program will be able to provide useful comparative effectiveness information. AHCPR views its budget as insufficient to permit sole funding of major RCTs, although the agency has on a few occasions collaborated with other agencies (e.g., the Veterans Administration (VA) and several institutes within the National Institutes of Health (NIH)) to take part in a larger comparative effectiveness study.

Traditionally, RCTs have been the tool associated with narrowly defined efficacy studies, and they have been justifiably criticized for their frequent lack of applicability to the broad range of patients and problems encountered by clinicians in everyday practice. However, RCTs need not be a narrow tool. Variations of the RCT design can be applied to comparisons among existing interventions, and to include broadly representative populations and settings. Examples of innovative and potentially useful approaches are:

- large, simple trials—trials with very simple protocols that enable research to include hundreds of thousands of participants and to be carried out in community practice settings: and
- trials that use innovative units of randomization--e.g., trials that randomize patients to different practices, or that randomize providers or geographic areas (instead of patients) in order to test different clinical management strategies.

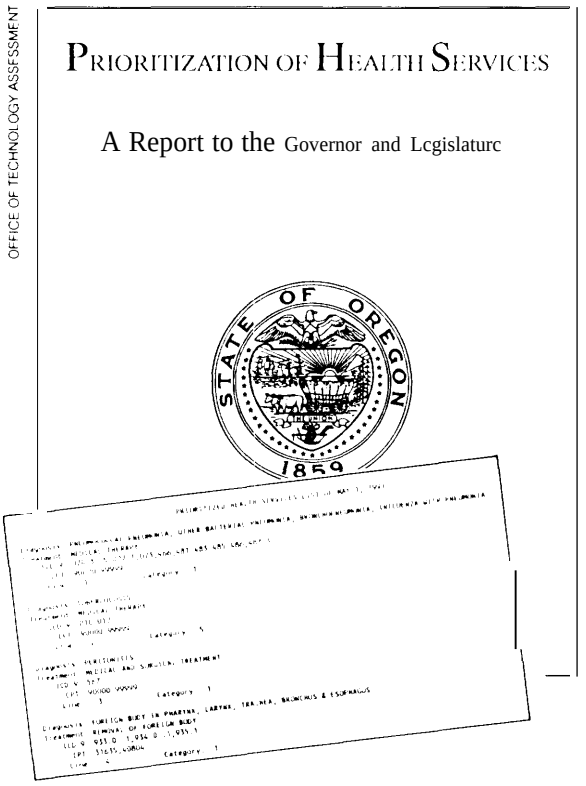
RCTs are especially important research design in studies where the differences in outcomes of the interventions being compared may be statistically modest but clinically important.

Interestingly. NIH. the premier federal sponsor of biomedical research, may well already conduct many clinical trials on medical technologies and practices that are in widespread use. Howeverr. that agency does not generally coordinate its clinical research resources with research questions generated by AHCPR. Nor are NIH's clinical trials documented in a way that makes it possible either to know how resources are being allocated in experiments of existing versus new technologies, or to critique the NIH clinical trials effort overall. Compiling an accurate and reasonably detailed database of NIH current activities. and assessing those activities, would greatly aid policy makers when contemplating changes in the federal investment in understanding the implications of current medical practices.

The National *Institutes of Health the federal governments premier sponsor of biomedical research and development, spends about 10 percent of its $10 billion budget on clinical* trials mostly *trials of new technologies* At *present there are few links be between N/H trials and AHCPR-generated research prlorities*

Cost-Effectiveness Analysis

Cost-effectiveness analysis (CEA) is a structured, comparative evaluation of two or more health care interventions. CEA can improve public and private policy makers' decisionmaking by structuring and making explicit the full range of costs and health effects relevant to a decision. Although CEA is still not routinely applied to most health care decisions, the sponsorship, use, and interest in these analyses have been increasing rapidly.



Not only health care providers but policymakers are increasingly interested in applying information on cost-effectiveness to their dec/s/ens, as evidenced by Oregon's attempt to use cost-effecfiveness as one criterlon for prlorltlzing healfh services under its Medicaid program

As the use of CEA increases, attention to the validity and comparability of analyses becomes crucial. Inconsistencies among analyses in the approaches and assumptions they use will confuse policy makers and hinder the practical use of CEA. U.S. and international efforts to address this issue, through better standardization of at least some aspects of CEA, deserve attention and support.

Cost-utility analysis (CUA) is a form of CEA in which quality-of-life outcomes of interventions being compared are incorporated quantitatively into the analysis (e.g., as "quality-adjusted life years," or "QALYs"). CUA is potentially attractive to policy makers because it facilitates comparisons across health care interventions with very different purposes. Because CUA incorporates some social preference factors directly into the analysis, however, users must be doubly careful to bear in mind that—like other forms of CEA—this technique cannot address, and may obscure, some of the most crucial social policy concerns.

The quantitative calculations in CUA, for example, do not allow for the fact that society is not always indifferent to which groups benefit and which do not; an intervention that looks the most positive when measured by cost per QALY may in fact not always be the "best" allocation of social resources when these concerns are taken into account. Nor does CUA address the question of whose values should matter the most for particular decisions; it treats all values as the social average. A third caution for users is that in applying CUA, one is assuming that the preferences for various states of health reported by people in surveys translate into accurate representations of their beliefs about the value of different interventions or resource allocations. This assumption has not been validated empirically.

Another very significant change in cost-effectiveness methodology is the growing practice of conducting CEAs simultaneously with early clinical trials of a new treatment efficacy and safety. Such studies may be biased towards finding no difference in costs between treatments, even where one exists, because the economic questions may require larger sample sizes to obtain statistically significant results than the health outcome questions. More fundamentally, these trials raise familiar issues of generalizability: the cost results derived from an efficacy trial may not be applicable outside of the trial, in ordinary practice.

Despite the concerns about their comparability and uses, cost-effectiveness studies and related activity in the private sector have boomed. Private industry, spurred by the need to deal with an increasingly sophisticated cadre of managed care administrators who are very cost conscious, has begun putting significant resources into efforts to show that its products are not only clinically effective but cost-effective. The pharmaceutical industry in particular has become very active in sponsoring cost-effectiveness analyses of its new

products. To the extent that the results of these analyses are used in marketing claims, both purchasers (e.g., government and private insurance programs) and regulators (i.e., the Food and Drug Administration) will need to become increasingly sophisticated at evaluating the claims.

Given the growing level of interest among private and public policy makers alike in CEA. the federal government's level of activity in this area is surprisingly weak. Only in the area of preventive services is there an y significant federal investment. CEA and supportive methodological research related to treatment and long-term management have been given relatively little attention by federal agencies. There is no uniform agreement about what role information about the cost-effectiveness of treatments should play in private or public insurance coverage decisions, but more agreement on this point may emerge in the near future. At present, federal agencies are not well-positioned to support CEA-related research, through either in-house expertise or current sponsorship of methodological studies.

Health Technology Assessment and Clinical Practice Guidelines

"Health technology assessment" as used in this report is a structured analysis of a health care technology, a set of related technologies, or a technology-related issue that is performed for the purpose of providing input to a policy decision. The federal role in health technology assessment has been an ongoing topic of debate since the field emerged in the 1970s. Recent changes, however. have given this debate a new twist.

One of the most remarkable developments in the field of health technology assessment has been the explosive growth in the private sector market for assessments of specific medical technologies. A few individual private-sector payers and providers have had some involvement in health technology assessment for years. What is new, however, is the degree to which technology assessments are becoming a standard ingredient in private-sector decisionmaking. This trend is likely to continue, in parallel with the growth in managed care.

Responding to this demand, the private market in health technology assessments has become a full-fledged economic activity in its own right. Many larger insurers and provider organizations have in-house staff dedicated to the endeavor. Others interested in assessments of particular technologies can now turn to private consulting firms, academic departments, and other organizations that have assembled the needed expertise and made their assessments widely available.

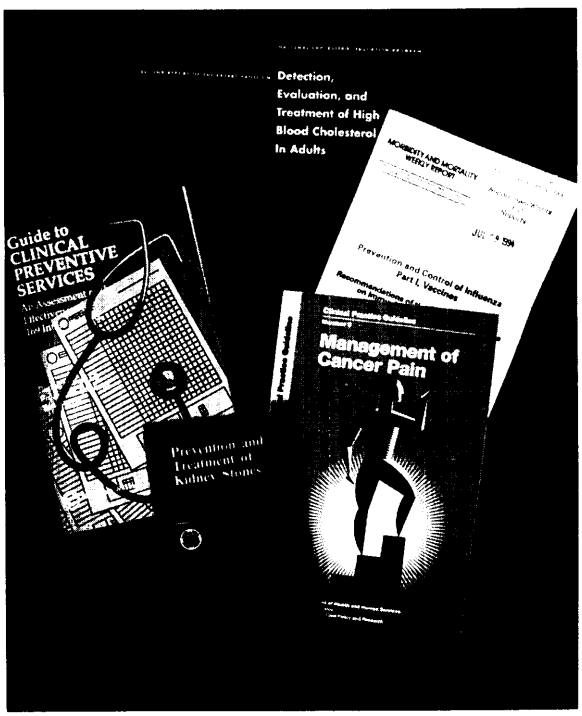
Meanwhile, the federal government's investment in assessments of individual technologies has been centered on the Office of Health Technology Assessment (OHTA), an office located within AHCPR that undertakes assessments of particular health technologies at the request of the Medicare and C HAM PUS [7] programs.[8] That office activities have been largely unchanged in degree over time.

While few federal agencies produce detailed staff assessments of individual technologies, many federal agencies sponsor and issue health technology assessments in the form of clinical practice guidelines. These agencies include AHCPR, the Centers for Disease Control and Prevention, several components of the National Institutes of Health, and the Office of Disease Prevention and Health Promotion (ODPHP). In all of these cases, the guidelines are developed by an expert panel sponsored by the agency, not by agency staff.

[7] CHAMPUS is the acronym for the Civilian Health and Medical Program of the Uniformed Services, operated by the Department of Defense for military retirees and dependents.

[8] The congressional Office of Technology Assessment also performs health technology assessments. but because it is located in the legislative branch of the government, It\ role in producing technology assessments i\ I imited to studies requested by Congress.

OFFICE OF TECHNOLOGY ASSESSMENT



Numerous federal agencies sponsor clinical practice guldelines, but their guideline development efforts are uncoordlnated and their recommendations often conflict with each *other*

Federal guideline development efforts are often considered to be philosophically distinct from other efforts to assess health care interventions. This distinction is especially notable in AHCPR, where the guidelines effort has more organizational and historical links with effectiveness research than with health technology assessment. The distinction, however. is an artificial one. Guideline development efforts are simply a different manifestation of the need to assess the impacts of health technologies. Even if guidelines are intended primarily for individual educational purposes, they constitute decisions about the best use of medical technologies that are implicitly supported by the federal government.

Clinical practice guidelines do have some unique attributes. In particular, unlike other federal technology assessments, they involve clinical experts or other public representatives of affected groups as the assessors themselves. The methods by which clinical practice guidelines are derived and the impact of those methods on the guideline recommendations for practice have received little attention. Given the prominence of guidelines as a component of many of the proposals to improve the health care system currently being debated, this deficit is very disturbing.

Federal guideline development is also hindered by a lack of coordination. There is no overall principle or strategy that guides the many guideline development efforts, and different agencies sometimes issue guidelines on the same topic. Although in some cases the recommendations of one agency are explicitly adopted by another, recommendations can conflict as well. Furthermore, recommendations from federally sponsored guidelines can conflict with guidelines on the same topic promoted by private groups. Differences among guidelines recommendations can cause confusion and may undermine the basic credibility of guidelines themselves.

The enormously varied methods used by the various private and federally sponsored groups to develop clinical practice guidelines contribute to conflicts in their recommendations. Examples of methodological differences include:

- the degree to which expert panels follow strictly specified formal rules of group interaction to arrive at consensus,
- the degree to which they rely on scientific evidence of benefit to support their recommendations,
- the diversity of experts represented on the panel, and
- the degree to which guidelines are explicitly structured to account for factors such as cost and patient preferences.

A few federally sponsored guidelines have included assessments of the guidelines' likely impact on health care costs. None, however, has explicitly laid out the comparative costs of alternative technologies or management strategies being considered in the guideline, and formally incorporated this analysis into the recommendation-making process.

Linking guideline recommendations to good evidence improves the validity of guidelines and the likelihood that panels of experts **will**

agree on practice recommendations. Evidence-based clinical practice guidelines have proved workable and politically acceptable. The number of organizations that base their guidelines on an explicit review of evidence, and formal methods of linking recommendations to the strength of that evidence, is small but growing. The theoretical strength of such guidelines at the national level is so compelling that it calls into question the usefulness of federally sponsored guidelines that are not evidence-based. Guidelines with less formal links between evidence and recommendations may be justified for some purposes (e.g., guidance on the use of very new technologies), but those purposes should be carefully thought out.

An advantage of linking recommendations clearly to existing evidence is that it can help identify high-priority research areas. Outlining clearly the most important clinically relevant questions for further research is an important contribution of guideline efforts that is often underemphasized.

Group composition and aspects of group process become increasingly important determinants of guideline recommendations as the availability of evidence declines. For example, whether or not panelists perform the procedure under consideration seems to affect group judgments. Guideline recommendations also are sensitive to aspects of the guideline process (e. g., definition of appropriateness). In general, formal group process techniques seem to improve group performance, but this has not yet been verified in the context of clinical guideline development.

It is important to establish which processes produce valid and usable guidelines. At present the various guidelines approaches vary markedly in terms of resource use, yet there is no clear indication as to whether one method produces a guideline that is any better than another. It may be that some processes are particular y appropriate to certain purposes or under certain circumstances, but at present there is little evidence upon which to tailor guideline efforts.

Changing Clinical Practice

Clinical decisions are shaped by the evidence of potential risks and benefits, the judgments of clinicians and patients about the relative desirability y of possible outcomes, and a range of external forces. External influences that can affect whether clinicians change their practice in response to clinical practice guidelines, or other sources of information, include the following:

- financial incentives, such as payment rates, bonuses, and salaries:
- administrative influences, including payment denial, utilization review, prior authorization requirements, and other mechanisms; and
- the advice of clinical colleagues, acknowledged clinical experts, and organizations with which the practitioner is associated.

Available studies and experience suggest that merely disseminating clinical practice guidelines will often be insufficient to change practice. Changes in practice are more likely if implementation efforts are more active and intensive; if they involve multiple- rather than single-pronged approaches: and if the efforts are tailored to specific context and problems addressed by the particular guideline. The ability to adapt guidelines to local circumstances may enhance their acceptance (but may also permit variations in practice to continue).

Physicians are more likely to ascribe credibility to information from sources they know and respect. Personal involvement in the process of change is also an element common to many successful efforts to alter practices. These features present a dilemma to government sponsors of guidelines, because guidelines developed by clinicians, and particularly clinical specialists, may not reflect the values of nonclinicians or nonspecialists who are also affected by the guidelines.

Financial and administrative mechanisms can be powerful agents of change, but they do have substantial limitations. They are insufficient tools to improve practice, because they do not them-

selves identify which choices are most likely to be effective. They are also often perceived as lacking credibility because they are usually externally generated, and they may have unintended results if clinicians attempt to circumvent the actions being promoted. Changes brought about through economic and administrative mechanisms may not be durable if the mechanisms are removed.

Some clinical practices are more amenable to change than others. Cancer screening practices, for example, can be increased using computer and manual reminders, as well as a variety of other administrative mechanisms. Guidelines for the use of x-rays, blood tests, and pharmaceuticals have also been implemented successfully. Interventions to change practice have been less successful for more complex clinical decisions, such as choosing between medical and surgical treatments, or managing complex medical problems.

High-quality evidence alone (e.g., evidence from RCTs) will not necessarily lead to changes in clinical practice. However, clinical practice guidelines supported by strong evidence are more likely than are other guidelines to effect changes through such mechanisms as utilization review, computerized protocols, opinion leader educational efforts, or economic incentives.

Data collected in the course of routine patient care and by health insurance companies are increasingly being used in efforts to change clinical practice. Collated provider data ("practice profiling" or "report cards") are used to promote discussions about correct practice among physician colleagues, to compare the outcomes of care across physicians and institutions as a means of targeting quality improvement efforts, and to compare patterns and costs of care so that payers and employers can choose providers or negotiate rates,

These applications do, at least under some circumstances, lead to changes in clinical practice. Without the benefit of "benchmarks" based on knowledge of the most effective practices or other evidence on the comparative effectiveness of different practices, however, these applications are unreliable and will not necessarily lead to better care. (If there is no basis for knowing which pattern of care is, on average, better, reducing variation may still reduce costs but may face more provider opposition.)

Because so many factors influence clinical decisionmaking, no single strategy for implementing clinical practice guidelines will be uniformly effective. Successful strategies will be intensive, intervene through several pathways, and be tailored to the particular clinical problem and task. Consequently. changing clinical practice will not necessarily be either cheap or easy. Additional research is needed to illuminate more clearly the forces and strategies that influence clinical decisionmaking, and to test strategies for changing the often complex decisions of practicing community physicians.

OPTIONS FOR ADVANCING THE FEDERAL EFFORT

The current federal effort to improve health care services through the evaluation of health care interventions is being carried out through a wide variety of agencies and departments (table 1-1). This effort is strongly hampered by gaps in the existing research effort, by uncertainties in the federal role for health technology assessment, and by duplication and lack of coordination of clinical practice guidelines development.

Options for Congress and federal agencies in addressing these problems are presented below. Options to address research needs are summarized briefly (see chapters 4 and 5 for in-depth discussions). options relating to federal technology assessments and clinical practice guidelines are presented in slightly more detail.

Filling the Gaps in Effectiveness and Cost-Effectiveness Research

The crucial question for the next stage ineffectiveness and cost-effectiveness research is how to address the gaps that currently exist in this research. Some of these needs, and options for addressing them, include:

TABLE 1-1: Federal Agencies That Evaluate Health Care Technologies[1]

Agency	Primary function	Relevant evaluation activities reviewed in this report
Department of Health and Human Services (DHHS)		
• Health Care Financing Administration	Administers the Medicaid and Medicare programs	▪ Effectiveness research ▪ CEA
Public Health Service		
• Agency for Health Care Policy and Research	Conducts, supports, and disseminates research on health services, health care costs, and the effectiveness of clinical practices,	▪ Effectiveness research ▪ CEA ▪ Assessment of individual technologies ▪ Clinical practice guidelines
▪ Centers for Disease Control and Prevention	Administers national programs for the prevention and control of communicable diseases and environmental problems	▪ Effectiveness research ▪ CEA ▪ Clinical practice guidelines
• National Institutes of Health	Conducts and supports biomedical research into the causes, prevention, and management of diseases	▪ Effectiveness research ▪ CEA ▪ Assessment of individual technologies ▪ Clinical practice guidelines
▪ Office of the Assistant Secretary for Health--Office of Disease Prevention and Health Promotion	Promotes health education, supports and coordinates prevention programs among agencies in DHHS	▪ CEA ▪ Clinical practice guidelines
Department of Veterans Affairs		
▪ Veterans Health Ad m in istration	Administers and coordinates the delivery of health care to veterans	▪ Effectiveness research[2]

[1] The Fooc and Drug Administration has a strong role in setting standards for evaluator of technologies but does not itself conduct research or assessments

[2] The Veterans Health Administration also produces guidelines and economic studies for Internal use but those efforts were not evaluated in this report

KEY CEA - cost effectiveness assessment DHHS = Department of Health and Human Services

SOURCE Off Ice of Technology Assessment, 1994

1. Improving the efficient production of meta-analyses and other systematic reviews of existing studies, to make the best use of past efforts at clinical evaluation.

Options:

. Increase funding targeted to systematic reviews (e.g., through specific grants, PORTS, or the U.S. participants in the Cochrane Col - laboration).[9]
- Require investigators proposing new clinical studies to demonstrate. through references to meta-analyses or other systematic reviews, that the research is not unnecessarily redundant.

[9] The Cochrane Collaboration is an international netw ork of researchers w 110 are committed to preparing, maintaining, and disseminating systematic reviews of clinical trials on health care topics (see chapter 4).

- Encourage the National Library of Medicine to maintain a commitment to establishing comprehensive databases of published controlled clinical trials.

2. Conducting more, and more efficient, clinical trials that yield valid comparative information on health technologies already in use, to produce results directly useful to patient and clinician decisionmaking; and making valid, well-designed comparative studies an intrinsic part of ordinary practice in every setting.

Options:

- Encourage collaboration among AHCPR and NIH researchers. particularly regarding the wider use of broad outcome measures in more NIH-sponsored clinical trials.
- Establish and maintain a comprehensive database of ongoing clinical trials sponsored by the federal government (and, where possible, private industry).
- Invest in a nationwide. community-based research infrastructure that could be used for conducting large, community-based clinical trials on topics of broad interest to practitioners and patients.

3. Encouraging greater comparative evaluations of newly introduced technologies.

Options:

- Offer incentives to manufacturers to conduct comparative effectiveness studies.
- Encourage or require payers, including government insurers, to link health insurance coverage for new technologies with structured, monitored evaluation of those technologies.
- Expand the federal government role in sponsoring comparative evaluations of new technologies.

4. Encouraging appropriate development of CEA. As the private sector becomes increasingly interested in producing and using cost-effectiveness analyses, both as an evaluation and a marketing tool, federal regulators and health care payers need to become educated users and reviewers of these analyses. Public policy makers, too, have a vested interested in access to high-quality, comparable CEAs as a tool for decisionmaking.

Options:

- Coordinate and underwrite efforts to improve the comparability of CEAs being produced in both the public and the private sector.
- Increase sponsorship of policy-relevant CEAs and of underlying methodological research, particularly research that examines the different impact of different methods on analytic results.

Clarifying the Federal Role in Health Technology Assessment

OHTA, a component of AHCPR, has recently been instructed by Congress to set priorities for technologies to assess in the event it can conduct some assessments for private-sector users (Public Law 102-41 O). Given the vastly expanded private sector capability for individual technology assessments, however, payers, providers, and others wanting assessments of particular technologies will often be able to obtain them elsewhere. Thus, the future role for government-sponsored assessments could take several possible paths.

Options:

- Focus OHTA efforts on the needs of federal payers.

  OHTA could expand the breadth of its assessments (e.g., to more technologies) and the breadth of government programs for which it perform assessments (e.g., Medicaid programs) but could continue to perform assessments only at the request of government payers or other decisionmakers. (Under health reform scenarios that include some form of future national health benefit-setting board or agency, OHTA, or its equivalent, might need to expand its capacity considerably.) Exceptions could be made for unusual circumstances in which an assessment is believed to be vitally needed and for some reason is not being conducted, or can-

not be adequately conducted, in the private sector. The advantages of this option include limiting tax-funded expenditures for individual technology assessments to those specifically needed by government programs. This option would also continue to permit the private sector to produce multiple assessments that could be compared, discussed, and targeted to the specific needs of the users that sponsor or purchase them.

- Alternatively, expand OHTA's capabilities to accommodate the needs of private sector users.

  (Under health reform scenarios in which detailed benefits are set at the regional or local level by private-sector plans, OHTA would perform assessments for these users. for exampie.) The advantages to this option relative to the previous one are the efficiency of a single source of assessments, so that private payers and providers are not faced with conflicting conclusions or duplicated efforts. and so that critiques of the assessments can be focused in a public forum. Potential disadvantages are greater government expense and less opportunity for multiple. targeted assessments. If this option were chosen, OHTA would need to greatly increase its size and scope to accommodate user needs.
- Increase OHTA sponsorship of privately produced assessments.

  Under either of the above alternatives, an intermediate course is possible under which OHTA sponsors technology assessments. or assessment centers, but many of those assessments are actual] y performed under contract or agreement by private assessment organizations.

Improving Clinical Practice Guideline Development

If clinical practice guidelines are to continue to be an important component of health reform strategies, the implications of how they are developed need to be explored and understood in much more detail.

Options:

- Develop better methods and clearer rationales for prioritizing guideline topics.

  Priorities for guideline topics may depend on the purpose of guidelines. If they are to be used as educational tools to improve adherence to effective practices. an important criterion for selecting guideline topics is whether sufficient evidence exists to form the basis for a credible and reliable guideline. A second criterion is whether actual practice varies from that expected based on the evidence. Variation in practice alone, however, is an insufficient reason to develop a national clinical practice guideline for this purpose.

  Other criteria for choosing topics might become more important for other purposes-e. g.. if the guidelines are to provide immediate information on the status of a very new technology; or if the guidelines are to establish which, among clinically acceptable management strategies, arc the least expensive strategies.
- Document and test alternative methods and models for guideline development.

  There is no solid basis at present for judging whether one method of developing guidelines is better than another. but neither are there really ongoing activities that will help future policymakers make such judgments. Existing group processes used by the guidelines panels themselves. particularly formal ones (e.g.. the Delphi approach used by some expert panels) could be further developed and tested and contrasted with one another.

  Little research has been done on the crucial areas of different methods to incorporate cost assessments and patient preferences into practice guidelines, and contrasting the effects of different methods on the guidelines' formats and recommendations.

  In addition, there are a number of possible alternative models for the federal role in guidelines development. For example, one alternative model to test would be to create standing (cams to support guideline panels. Such teams might perform several of the more technical or

less clinical tasks that guideline panels must do: conduct literature reviews, assess current practices, and perform cost or cost-effectiveness analyses. The expert panel that ultimately developed the final guideline recommendations under this model might be a federally sponsored panel, or it might be regional or local health plans or providers.

- Strengthen the federal investment in the development of tools that can be used by guidelines panels, public and private alike.

  Federal agencies are in the unique position to be able to assemble resources needed for guideline development. Some of the tools to enhance the efficiency, reliability, and credibility of future guidelines are those that would fill in some of the gaps of effectiveness research, including comprehensive databases of clinical trials and support for systematic reviews of topics of interest. Other useful areas in which tools could be developed include developing additional sources to identify areas of clinical uncertain y (e.g., national databases to identify practice variation; national clinician surveys and focus groups to assess sources of variation).

- Coordinate guideline efforts across agencies.

  At present, the potential for unnecessary duplication and contradiction between guidelines, and inefficient cross-agency use of resources needed to produce guidelines, is high. Only for prevention guidelines does some structure to address this problem nominally exist, through the Office of Disease Prevention and Health Promotion. However, ODPHP has no jurisdiction over treatment or long-term clinical management guidelines, the categories in which NIH and AHCPR are most likely to overlap.

Directing and Coordinating the Overall Federal Effort

Filling the gaps in the federal effort to evaluate health technologies in current use will require greater coordination among agencies. It will also require either new resources or shifts in the priorities and purposes to which existing resources are committed.

Most difficult of all, filling the gaps in the federal effort will require changes in the perceived responsibilities of several agencies, particularly AHCPR and NIH. Although AHCPR is at the moment the designated focal point for the federal effectiveness initiative, that agency does not currently have the mandate, the commitment, the resources, or the leverage either to fill the gaps entirely itself or to successfully coordinate the effectiveness research and clinical practice guideline efforts of other agencies.

Options:

- Designate a single lead agency to perform effectiveness research activities and coordinate guideline activities.

  Alternative strategies for achieving this centralization would be to fold AHCPR into a new, larger agency with a broader mandate and more resources; or to change AHCPR's mandate (or the mandate of another agency) to designate that agency it as the lead agency for coordinating guideline efforts, for conducting comparative effectiveness trials, and for filling some of the other most pressing needs.

- Do not establish a single lead agency, but clarify the roles of existing agencies in effectiveness research and encourage or require collaboration among agencies through administrative mechanisms.

  For example, a possible mechanism for collaboration might be to require NIH institutes to give high priority to funding research studies on topics identified by guideline panels, PORT findings, or advisory bodies at AHCPR.

The great advantage of designating a single, larger agency as the focal point to fill the gaps in effectiveness research is that coordination across agencies is inevitably cumbersome, time-consuming, and haphazard in many ways. However, this strategy also has substantial disadvantages, including:

- the problem of causing fresh organizational disruption only six years after the creation of AHCPR:
- the difficulty of any single agency actually encompassing all relevant activities (e.g., all clinical trials on existing therapies. or all clinical practice guideline development, including those current] y sponsored under the auspices of NIH institutes and CDC):
- the difficulty in finding additional funding to expand these activities; and
- the danger that, without substantial additional resources, any new agency will be unable to im - provc significantly on the current commitments of AHCPR.

Clarifying and respecifying the roles of existing agencies to fill the gaps in effectiveness research is a much less expensive and, in some ways, a simpler strategy. Implementing this opt ion, however, would require a shift in funding between or within agencies towards studies performing comparative research on existing practices and technologies, rather than towards the development of new technologies or descriptive studies. The organizational and institutional barriers to shifting either resources or research priorities are themselves substantial and would probably require a legislative directive to overcome.