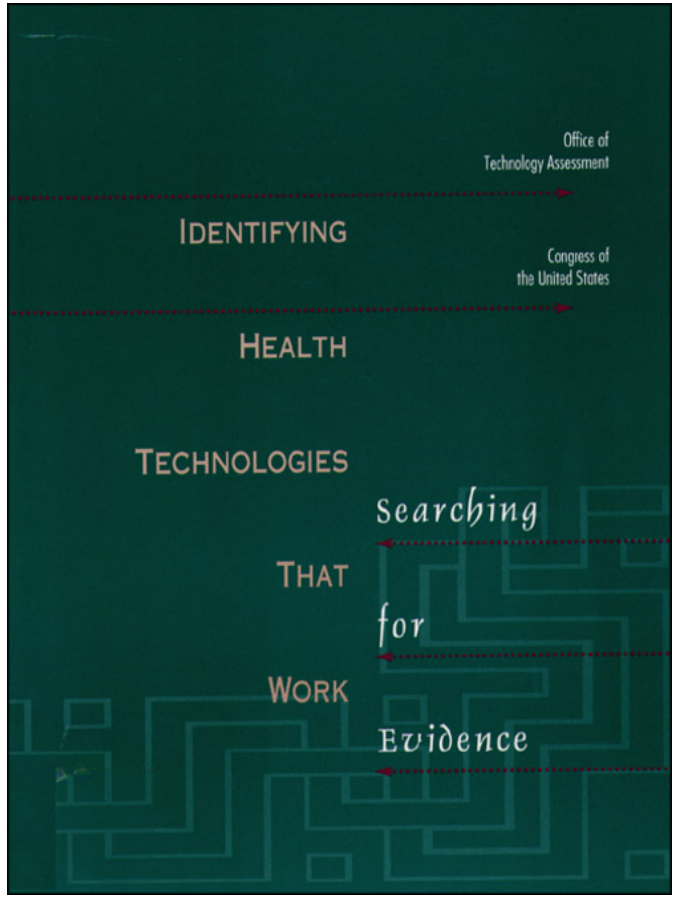
*Identifying Health Technologies That Work: Searching for Evidence*

September 1994

OTA-H-608
NTIS order #PB95-109617
GPO stock #052-003-01389-4



**Recommended citation:** U.S. Congress, Office of Technology Assessment, *Identifying Health Technologies That Work: Searching for Evidence, OTA-H-608* (Washington, DC: U.S. Government Printing Office, September 1994).

For sale by the U.S. Government Printing ot'flee
Superintendent of Document>, Mail Stop: SSOP, Washington. DC 20402 2-9328
ISBN 0-16 -045234-1

# Foreword

"Clinical practice guidelines". . . "health plan report cards". . . *'outcomes research". . . '-inappropriate care''—all of these are popular phrases in the current debate about how to reform the American health care system. Each, it is hoped, will be a contributor towards creating a more comprehensive system that will still be able to offer high quality. moderate cost care. Underlying each, however, is a single basic assumption—that researchers can accurately identify which health technologies work better than others, and that clinicians and patients will use this information in everyday practice.

The federal government is the main sponsor of research to evaluate health technologies currently in use. The purpose of this report is to examine two crucial questions:

1. What are we getting out of this investment?
2. How can we improve it?

The idea of studying health technologies to distinguish effective from less effective, and less cost-effective, technologies is a longstanding one. Many of the techniques now being applied, however, are new, or have new applications, or have received a new emphasis in recent years. Accordingly, these techniques offer fresh opportunities, but they also come with new caveats about their use. By understanding both the possibilities and limitations of current methods of evaluating health technologies, and addressing deficiencies in the federal enterprise, we can take the next step in the effort to identify health technologies that work.

This assessment was prepared in response to a request by the Senate Committee on Labor and Human Resources. The report was prepared by OTA staff, with assistance from several researchers who prepared background papers under contract to OTA describing specific research techniques in detail. (The collected set of background papers is available separately. ) In addition, OTA gratefully acknowledges the contributions of the assessment's advisory panel and many other individuals who provided valuable information and reviewed preliminary drafts. As with all OTA documents, the final responsibility for the content of the assessment rests with OTA,



ROGER C. HERDMAN
Director

# Advisory Panel

**William Fullerton, Panel Chair**
Crystal River, FL

**Robert Brook**
Director
Health Sciences Program
The RAND Corp.
Santa Monica, CA

**Iain Chalmers**
Director
The Cochrane Centre
Oxford, England

**Harold Cohen**
President
Hal Cohen, Inc.
Baltimore, MD

**David Eddy**
Professor
Health Policy and Management
Duke University
Durham, NC

**Ruth Faden**
Professor and Director
Program in Law, Ethics, and Health
Johns Hopkins University
Baltimore. MD

**Diana Jest**
Vice President, Government Affairs
Group Health Association of America
Washington, DC

**Robert Keller**
Executive Director
Maine Medical Assessment Foundation
Augusta, ME

**Jonathan Lomas**
Professor
Centre for Health Economics and Policy Analysis
McMaster University
Hamilton, Ontario

**Michael McCulley**
Assistant General Counsel
Johnson & Johnson
New Brunswick, NJ

**Barbara McNeil**
Ridley Watts Professor and Head
Department of Health Care Policy
Harvard Medical School
Boston, MA

**Frederick Mosteller**
Director
Technology Assessment Program
Harvard School of Public Health
Boston. MA

**Richard Peto**
Clinical Trial Service Unit
Oxford, England

**John Wennberg**
Director
Center for Evaluative Clinical Sciences
Dartmouth Medical School
Hanover, NH

**Note:** OTA appreciates and is grateful for the valuable assistance and thoughtful critiques pro} ided by the advisory panel members. The panel does not, however, necessarily approve, disapprove, or endorse this report. OTA assumes full responsibility for the report and the accuracy of its contents.

# Project Staff

**Clyde J. Behney**
Assistant Director, OTA

**Sean R. Tunis**[1]
Health Program Director

**Kerry B. Kemp**
Managing Editor

**ADMINISTRATIVE STAFF**

**Beckie Erickson**
Office Administrator

**Daniel B. Carson**
P.C. Specialist

**Carolyn Swarm**
P.C. Specialist

**Carolyn Martin**
Word Processing Specialist

**PRINCIPAL STAFF**

**ELAINE J. POWER**
Project Director

**Maria Hewitt**
Senior Analyst

**Cheryl Liechty**
Research Assistant

**Dorit Opher**
Research Analyst

**Sean R. Tunis**
Senior Analyst

**OTHER CONTRIBUTING STAFF**

**Michael Bee**
Analyst

**Alex Ching**
Intern

**Jill Eden**
Senior Analyst

**Laura Gottsman**
Research Analyst

**Arna Lane**
Research Analyst

**Leah Wolfe**
Analyst

**CONTRACTORS**

**Julie Buring, Charles Hennekens, and Michael Jonas**
Harvard School of Public Health

**Floyd J. Fowler**
University of Massachusetts at Boston

**Matthew Longnecker**
University of California at Los Angeles

**Neil Powe**
Johns Hopkins University

**Craig Tanio**
University of Pennsylvania

**Jeffrey Whittle**
Veterans Administration Medical Center—Pittsburgh

[1]Effective January 1994.

# Contents


## 1 Summary 1

Summary of Findings and Conclusions 3
- Effectiveness Research 4
- Cost-Effectiveness Analysis 7
- Health Technology Assessment and Clinical Practice Guidelines 9
- Changing Clinical Practice 11

Options for Advancing the Federal Effort 12
- Filling the Gaps in Effectiveness and Cost-Effectiveness Research 12
- Clarifying the Federal Role in Health Technology Assessment 14
- Improving Clinical Practice Guideline Development 15
- Directing and Coordinating the Overall Federal Effort 16

## 2 Behind the Search for Evidence 19

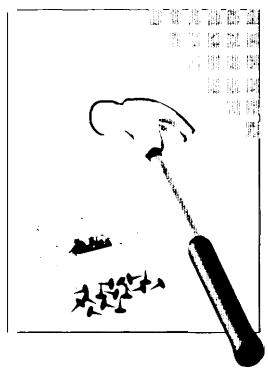
The Need for Evidence 20
A Framework for Evaluation 22
The Shaping of Effectiveness Research 26
- Geographic Variation in Medical Practice 26
- Outcomes of Patient Care 28
- Appropriateness of Care 30
- The Federal Medical Treatment Effectiveness Program 33

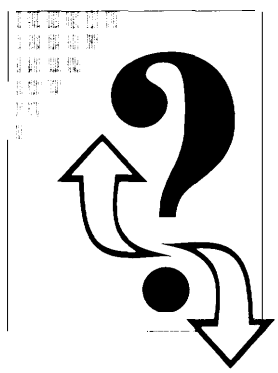
Expectations in the Context of National Health Reform 34
Conclusions 37

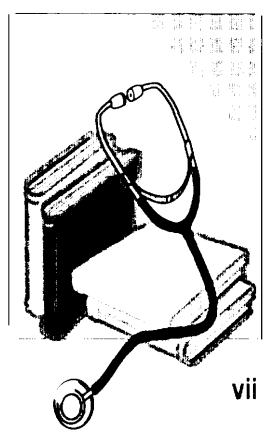
## 3 Tools for Effectiveness Research 39

Techniques To Evaluate Health Effects 40
- Basic Tools for Measuring Health Outcomes 40
- Primary Studies To Evaluate Health Effects 43
- Secondary Techniques To Synthesize Results 59

Applying the Tools: The Patient Outcomes Research Teams 67





Contributions 68
Limitations and Frustrations 72
Future Plans 73
Conclusions 73

**4 Issues in Improving Effectiveness Research 77**

Gaps in the Existing Federal Research Effort 77
Systematic Reviews: Making Use of Existing Knowledge 78
Filling In the Knowledge Gaps 82
The Comparative Evaluation of New Technologies 90
Issues in Federal Funding and Support 95
The Roles of the Federal Agencies 95
Coordinating Research Activities 102
Conclusions 104

**5 The State of Cost-Effectiveness Analysis 107**

Analytic Approaches 107
Uses of Cost-Effectiveness Analysis 108
Mechanics of Cost-Effectiveness Analysis 110
Issues in the Use of Cost-Effectiveness Analyses 111
Comparability 112
Quality of Analyses 114
Prospective Analyses: Clinical-Economic Trials 116
Underlying Assumptions of Cost-Utility Analysis 118
Federal Cost-Effectiveness Activities 122
Agency for Health Care Policy and Research 122
National Institutes of Health 122
Centers for Disease Control and Prevention 123
Health Care Financing Administration 124
Office of Disease Prevention and Health Promotion 125
Cost-Effectiveness Analysis in the Private Sector 125
Recent Trends 125
Issues 126
Conclusions 127

## 6 The Federal Role in Health Technology Assessment 131

Growth of Health Technology Assessment 132
- Debating the Federal Role 132
- Health Technology Assessment in the Private Sector 135
- Clinical Practice Guidelines and Technology Assessment 139

Conclusions 141

## 7 The Development of Clinical Practice Guidelines 145

Federal Guideline Efforts 145
Private Efforts 146
Guideline Issues Related to Development 147
- Overview 147
- Choosing Guideline Topics 149
- Selecting Guideline Panels 151
- Defining the Scope of Guidelines 153
- Identifying and Synthesizing Evidence 155
- Techniques To Aid Group Interaction and Decisionmaking 157
- The Basis for Guideline Recommendations 164
- Organization and Administration of Guideline Activities 166
- Methodological Research 168

Conclusions 168
- The Link Between Methods and Recommendations 168
- Developing Methods To Prioritize Guideline Topics 169
- Research Needs 170

## 8 The Impact of Clinical Guidelines on Practice 173

Anatomy of Physician Behavior 173
- Modeling Clinical Decisionmaking 173
- Sources of Influence on Physician Practice 175

Approaches To Changing Clinical Practice 177
- Passive Dissemination 177
- Active Approaches 179
- Lessons from the Literature 187

Implementing Clinical Practice Guidelines 190
- Impact of Guidelines on Clinical Practice 190
- Factors Influencing the Impact of Guidelines 195

Conclusions 197

APPENDIXES

A How This Study Was Conducted 199

B Acknowledgments 203

C Guideline Development Activities 210

D Abbreviations and Glossary 237

References 247

Index 295

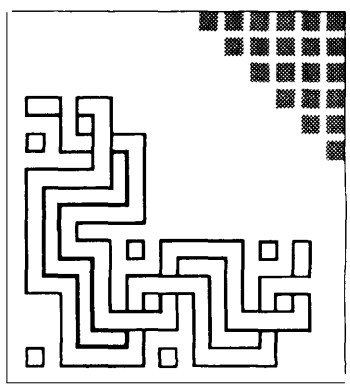
# Summary 1

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.

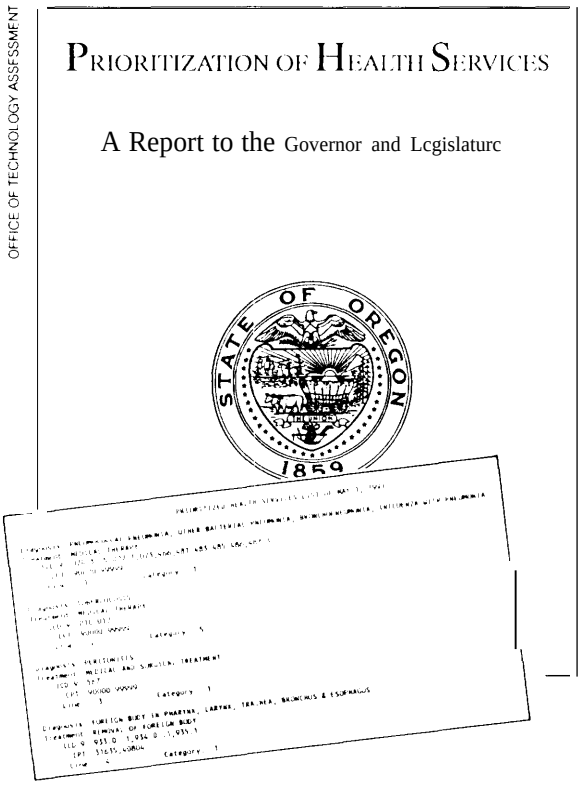
Interestingly. NIH. the premier federal sponsor of biomedical research, may well already conduct many clinical trials on medical technologies and practices that are in widespread use. Howeverr. that agency does not generally coordinate its clinical research resources with research questions generated 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.

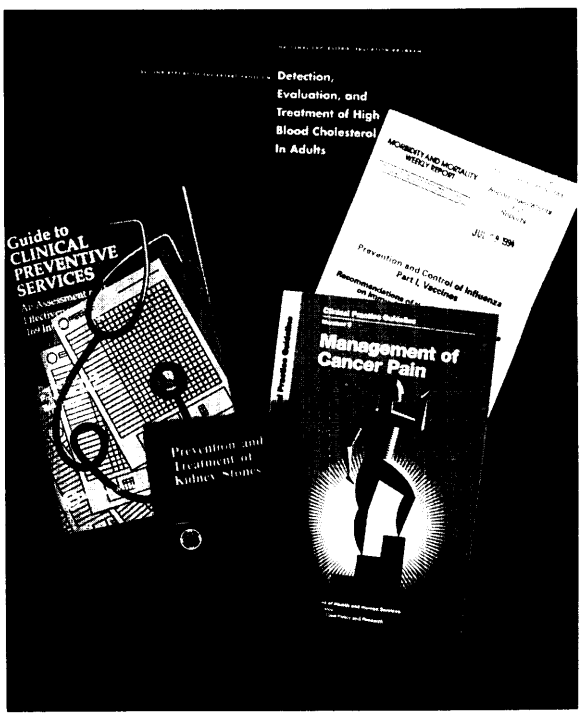
Meanwhile, the federal government's investment in assessments of individual technologies has been centered on the Office of Health Technology Assessment (OHTA ), an office located within AHCPR that undertakes assessments of particular health technologies at the request of the Medicare and C HAM PUS [7] 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<br>▪ 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<br>▪ CEA<br>▪ Assessment of individual technologies<br>▪ 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<br>▪ CEA<br>▪ Clinical practice guidelines |
| • National Institutes of Health | Conducts and supports biomedical research into the causes, prevention, and management of diseases | ▪ Effectiveness research<br>▪ CEA<br>▪ Assessment of individual technologies<br>▪ 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<br>▪ 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:

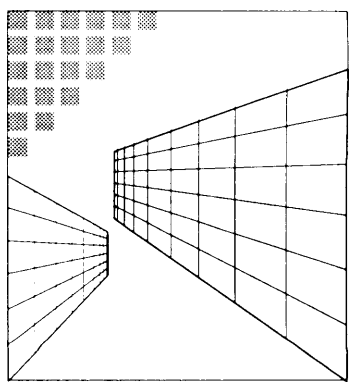
- the problem of causing fresh organizational disruption only six years after the creation of AHCPR:
- the difficulty of any single agency actually encompassing all relevant activities (e.g., all clinical trials on existing therapies. or all clinical practice guideline development, including those current] 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.

# Behind the Search for Evidence 2

The present interest in evaluating the worth of health technologies and clinical practices owes much to two men whose works were separated by time and geography. One, Earnest A. Codman, was a Boston surgeon practicing in the early 1900s. He believed adamantly that the path to improvement in medical care depended on documenting the outcomes of patients who had been treated. Codman's call for detailed public reports of these outcomes, including long-term followup assessments of patients, faced strong opposition in his own day and was never adopted on a large scale (562). Nonetheless, his work set the stage for modern day efforts to focus on comparative patient health outcomes as a basis for improving the quality and effectiveness of care.

Six decades later and a continent away, Archie Cochrane, a physician and epidemiologist. changed the way researchers, poli - cymakers, and clinicians viewed medical care with the 1972 publication of his book *Effectiveness and Efficiency: Random Reflections on Health Services (* 130). In it, he argued that to provide the best health care at a given level of national health expenditures, society first must improve the effectiveness (eliminating ineffective care) and efficiency ("the optimum use of personnel and materials") of the health care system (130).



Cochrane pointed out that a major cause of ineffective care was that too much medical decisionmaking was based on poor evidence—''expert opinion" or, at best. observational studies that could not adequately differentiate effective from ineffective (or harmful) medical care. He advocated an emphasis on randomized controlled trials to evaluate medical interventions. Most importantly, however, he stressed that more valid information on the effectiveness and cost-effectiveness of health care interventions

**BOX 2-1: The Tragedy of DES**

Diethylstilbestrol (DES) is a dramatic example of a drug that became widely used in clinical practice before it was found to have major adverse effects An estimated three million American women took DES between 1948 and 1970 (728).

DES became popular in the early 1950s, after the publication of several studies that suggested that it was efficacious In treating placental insufficiency, a condition that often causes stillbirths (1 54, 279,595,653,707a,927). None of these studies were randomized controlled trials and none used double-blinding. Five other contemporary studies that did use double-blinding failed to show that DES improved pregnancy outcomes (1 45,1 75,233,642,838). Nonetheless, individual cases of women who had had previous stillbirths, and who were finally able to have children after taking DES, provided physicians with anecdotal evidence supporting the management of high-risk pregnancies using this drug Despite the absence of reliable evidence supporting the use of DES, U.S. clinicians began to prescribe the drug widely

In 1970, two researchers published a paper that reported a number of cases of a rare cancer in daughters of women who took DES (338) A second paper published the following year found maternal usage of DES to be strongly associated with the development of tumors in young women (340). In 1971, the FDA announced that DES was contraindicated for use in pregnant women By this time, however, several million men and women had already been exposed to DES in utero Numerous studies have since identified a range of adverse effects, including increased incidence of certain rare cancers in DES children, reproductive system anomalies in both sexes, and an increased incidence of negative pregnancy outcomes for DES-exposed women (61 ,62,337,339,685,886).

The tragedy of DES is not only that the drug proved to be harmful to the children of women who took it, but it was never really shown to be effective even for the condition for which it was so enthusiastically prescribed. Ironically, a reanalyses of data used in one of the first studies that purported to support the use of DES found the drug to be associated with an increase in "miscarriages, 'premature' deliveries and neonatal deaths" (77)

SOURCE Adapted from PA Goldstein H S Sacks, and T C Chalmers Hormone Administration for the Maintenance of Pregnancy, I Chalmers M Enkin, M Keirse (eds ) *Effective Care in Pregnancy and Childbirth, Vol 7 Pregnancy* (New York NY Oxford University Press, 1992)

was crucial to improving both the quality and the efficiency of medical care.

## THE NEED FOR EVIDENCE

The basic foundation of the evaluation of a health technology (or any health care intervention) is information about its efficacy and safety: whether, under at least some conditions, the technology provides a health benefit that outweighs any attendant risks (779). The evaluation of efficacy and safety is far from a theoretical concern. Experience with technologies such as diethylstilbestrol (DES), a cancer-causing drug prescribed to millions of pregnant women in the 1950s and 1960s, has taught that even the most enthusiastically adopted technologies can be not only ineffective but lethal (box 2-1 ).

The federal government has long had a role in evaluating the efficacy and safety of certain categories of medical technologies. Within the Department of Health and Human Services, for example. the National Institutes of Health (NIH) conducts and sponsors both basic biomedical research and clinical trials to test some of the most

promising technologies developed by its scientists. The Food and Drug Administration (FDA) regulates drugs, biologics, and medical devices, requiring manufacturers to provide evidence of safety and efficacy before their products can be marketed.[1] Other departments such as the Department of Veterans Affairs (VA) and the Department of Defense often sponsor both the development and testing of technologies intended to improve the health of the population in their charge.

For all of this regulation and testing, however, society's understanding of the full effects of most of the health technologies it uses is remarkably small. This state of affairs has four causes.

First, much of what medical care has to offer was part of customary practice before rigorous testing for efficacy became common. Randomized, controlled trials to demonstrate the efficacy of interventions have been openly advocated only since the 1940s, and they have been used widely only since the 1970s (784). Yet drugs to treat open-angle glaucoma, for example, have been prescribed since the 1800s (400). The first randomized controlled trials of the effect of a drug in preventing vision loss due to open-angle glaucoma were not undertaken until the 1980s (227, 416).[2]

Second, a high proportion of newly introduced technologies, even today, are not required to show rigorous evidence of efficacy before they are adopted. Only the most novel medical devices, for example, are subject to individual scrutiny and approval by the FDA before they can be marketed (370). Therapies such as psychological counseling and surgical procedures are subject to no regulatory requirements regarding efficacy at all (except to the extent that they involve drugs or devices that are regulated). Promising new procedures thus are often widely publicized and adopted by physicians and patients without undergoing any formal evaluation (box 2-2). The longstanding estimate that only about 10 to 20 percent of procedures have ever been formally evaluated for safety and efficacy (924) remains a rule of thumb (e.g., see reference 208).

The third reason is that a technology, once introduced, is frequently used in circumstances that are quite different from those in which it was first shown to be efficacious. The effects of the technologies under the new conditions can be very different as well. Drugs tested and approved for use for one type of cancer, for example, are frequently used to treat other cancers as well (881). Neither providers nor patients can be certain that a treatment used for a new population or in a new setting will actually have the same risks and benefits as those shown in the initial efficacy studies.

And fourth, as meager as society's knowledge of the health effects of many medical technologies is, our knowledge of their economic and social effects pales by comparison. In 1982, the Office of Technology Assessment (OTA) concluded that "No class of technologies is adequately evaluated for either cost-effectiveness or social and ethical implications" (783). Recent observers have suggested that this is still the case (606).

Thus, the deficits in evidence regarding the value of existing health care interventions are substantial. Nonetheless, in the two decades since the publication of Cochrane's seminal work, the

[1] FDA regulations actually specify that the agency consider safety and "effectiveness," but FDA's interpretation of "effectiveness" is more akin to "efficacy" as used in this report. The kinds of requirements that medical products must meet to satisfy this standard depend on the type of product. Drugs, and some medical devices considered to present a high possible risk of harm to users, must meet the most stringent requirements. Medical devices in the lowest risk category are required to demonstrate only such features as whether the manufacturer of the device met standards for good manufacturing practices.

[2] New antiglaucoma drugs seeking FDA approval (e.g., topical timolol) have had to undergo rigorous testing for some time, but such drugs have had to show only that they could reduce intraocular pressure. Although high intraocular pressure is strongly associated with open-angle glaucoma, until recently no rigorous studies had actually investigated whether reducing intraocular pressure through drug therapy protected patients from losing vision (786a). The National Eye Institute is currently funding a large multicenter trial to examine more specifically which patients with slightly raised intraocular pressures would benefit from the preventive application of antiglaucoma drugs (853).

**BOX 2-2: Promising but Untested: Examples of Two Proposed New Surgical Therapies**

Surgical innovations are especially likely to enter mainstream medical practice without ever being exposed to formal testing The lack of tradition among surgeons in testing new therapies through randomized trials, and the perceived difficulty in conducting such trials, may explain some of this phenomenon In addition, however, new surgical therapies often are incremental, have theoretical appeal, and are not subject to regulatory oversight. All of these characteristics make surgical Improvements difficult to Identify and study before they diffuse into the health care system,

A recent example of a surgical innovation is a technique to improve lung functioning in emphysema patients (9) Emphysema is a potentially fatal disease in which extensive damage to lung tissue (usually as a consequence of smoking) Impairs respiratory functioning,

The new technique revolves the surgical removal of 20 to 30 percent of a patient's lung A similar technique was Introduced in the 1950s but was rejected by the medical establishment on the grounds that the removal of lung tissue to treat symptoms (i e., shortness of breath) caused by tissue damage could only have a negative impact on patients. However, the newly refined procedure has been tried in 20 patients, all of whom have reportedly shown functional improvements as a result of surgery No randomized studies have been performed to confirm that the apparent short-term improvements are real, and the long-term effects of the procedure remain unknown, If the technique captures the interest of physicians and patients it may never undergo further testing before being adopted into clinical practice, since it is not subject to the safety or efficacy standards of any regulatory body

Another example of the kind of innovation that may never undergo rigorous evaluation is a potential new surgical procedure to preserve the salivary glands of head or neck cancer patients, These glands are often destroyed during radiation therapy (80), To avoid such damage, a researcher at the Tufts University School of Dental Medicine has proposed transplanting the glands temporarily to the patient's abdomen After the last radiation treatment, the glands could be re-transplanted into the mouth

So far, the procedure has been attempted only in animals The biggest challenge facing scientists is making sure that the glands can survive long enough in their temporary location to enable a full cycle of cancer therapy to be completed However, this problem may soon be solved, Because this new procedure has considerable theoretical appeal, it could well become an accepted strategy in cancer care based primarily on a demonstration of its feasibility

SOURCE Off Ice of Technology Assessment 1954 based on sources as shown Full citations are at the end of the report

movement to improve the assessment of the health, economic, and social effects of health care technologies has increasingly, though erratically, gained momentum. One result of this movement has been the growing accumulation of "research-based evidence" (705). That evidence, in turn, can be used to support judgments about the value of the myriad components of health care: "evidence-based medicine" (3 15).

## A FRAMEWORK FOR EVALUATION

Improving medical care through increased knowledge about what works, and the applicat ion of that

knowledge, is a powerful concept. As support for the concept has increased, however, the language describing it has become increasingly muddled.

One of the most common phrases used to describe this effort is "outcomes research." The term originally arose to describe the line of health services research that has emphasized how little is often known about the effectiveness and outcomes of care that patients receive. This line of research, described in more detail below, ultimately led to the federal government's medical effectiveness initiative and the creation of the Agency for Health Care Policy and Research (AHCPR) to carry out this effort (Public Law 101 -239). The term has come to be used so sweepingly, however, that it has become problematic. For example, it is now often used synonymously with "outcomes-based management," a technique through which purchasers and providers hope to be able to manage the quality and cost of care provided to patients. This technique uses information on the outcomes of patients treated by a particular provider, or enrolled in a particular health plan, to stimulate actions that will improve care (box 2-3). The phrase "outcomes research" is rarely used in this report.

The convergence of terms has led to confusion among policy makers and the public alike between activities to improve the quality of care and those primarily aimed at identifying and improving its effectiveness. Although the concepts of quality and effectiveness are closely related—both are aimed at making health care "work" better—they are not identical. Activities to improve quality generally focus on improving the process by which an activity is performed, or the capabilities of those performing it, in order to improve outcomes. In contrast, research to investigate effectiveness focuses on what outcomes are associated with a given technology (or clinical management strategy, or any other health care intervention), and whether and under what circumstances that technology is better than alternatives. The relative effectiveness of a technology does indeed depend in part on how well providers perform it. Policy interventions to address problems in the quality of care, however, may be different from those interventions that address the overall effectiveness of care. The focus of this report is on the latter.

In this report, the phrase "effectiveness research" describes the category of research efforts aimed at identifying effective care and developing and refining methods to support the identification of effective care. The concept of effectiveness includes both whether the technology has a given effect and whether the technology is more effective than alternatives.

It is sometimes useful to make a conceptual distinction between efficacy and effectiveness. One generally wants to know whether a technology works at least under ideal circumstances (efficacy) before applying it more broadly (effectiveness).[3] In reality, however, the distinction between efficacy and effectiveness is often fuzzy. If the patient population in an initial efficacy study is sufficiently broad, for example, the study results may be credible evidence of effectiveness more generally. Conversely, a demonstration that a technology is generally effective in one population (e.g., women) does not necessarily imply effectiveness in a differently defined population (e.g., all adults).

Cost-effectiveness analyses are an increasingly common step in evaluating medical care. They use the results of effectiveness research, in conjunction with detailed cost information, as part of a structured, comparative evaluation of the relative costs and effects of two or more health care interventions.

Information on effectiveness and on cost-effectiveness, in turn, can form the basis of a health technology assessment: an analysis of a technology-related issue conducted for the purpose of

[3] For a more detailed discussion of usage of the terms "efficacy" and "effectiveness," *see the* OTA report, *Assessing the Efficacy and Safety of Medical Technologies (779).*

## BOX 2-3: Using Patient Outcomes in Health Care Management

Stimulated in part by research emphasizing the final health outcomes of patients as an endpoint for assessing care, health care payers and providers have become increasingly interested in "outcomes-based management " In this case, data on patient outcomes is used as a way to permit payers, providers, or patients themselves to make choices or implement programs that are hoped to improve the quality and cost of care

Integral to many of these efforts is some form of "report card, " a profile of data on the outcomes of patients treated by particular hospitals or physicians, or enrolled in particular health insurance plans. Among the measures of quality commonly found in report cards are mortality, rehospitalization, length of stay, childhood immunization rates, and cancer screening rates.

States and private organizations have been particularly active in embracing the use of report cards as an approach to quality monitoring and quality improvement. In some cases, the dissemination of cost and outcomes Information has been mandated by state governments (e.g , in Illinois, Missouri, and Pennsylvania) In 1988, New York State began collecting cardiac surgery outcomes data intended only for use by hospitals and physicians but was later forced to make the data public as the result of a lawsuit (721).

However, many providers have independently initiated report card programs to market health plans and as a means of identifying aspects of clinical management that deserve closer scrutiny Examples of such private sector activity include United HealthCare Corp , a large managed care network that has used quality indicators since 1991 (334,449,941 ); and the Cleveland Health Quality Choice Project, which in 1993 released its first assessment of the quality and efficiency of 31 participating hospitals in northeastern Ohio (364). Interested persons and organizations can purchase the project's report cards for a fee. The Maryland Quality Indicator Project, which was initiated in 1985 by the Maryland Hospital Association, now covers over 600 participating hospitals Among the 15 quality indicators measured quarterly are hospital-acquired infections, Cesarean sections, and unplanned readmission The data allow participating hospitals to compare themselves with their peers and decide what, if any, action to take in response to their results (271).

While the diversity in approaches to quality assurance indicates that such projects have a promising future in many environments, the variability has also meant that the field of quality measurement has remained largely unstandardized, confounding purchasers' ability to make meaningful comparisons among competing insurers or hospitals. Two recent nationwide projects have be-

providing input to a policy decision. In this latter case, the policy decision itself has ramifications for clinical decisionmaking.

Thus, the findings from effectiveness research may be applied directly by the practitioner and the patient to improve clinical decisionmaking. Alternatively, information on effectiveness may form part of the evidence base for more detailed analyses that incorporate information on costs and on other important social considerations. In the latter case, information on a technology's effectiveness affects clinical decisions and patient outcomes indirectly, by way of their incorporation into cost-effectiveness analyses, technology assessments, and policy decisions (figure 2-1 ).

Clinical practice guidelines created by expert groups lie in an intermediate area in this framework. They are sometimes treated as an extension

**BOX 2-3 continued: Using Patient Outcomes in Health Care Management**

gun to address this problem The most prominent quality initative is the development of a prototypical standard report card by a nonprofit organization called the National Committee for Quality Assurance (NCQA) Among the 21 managed care organizations participating in the effort are Kaiser Permanence and U S Healthcare, Inc (48), United HealthCare Corp (described above) is also participating in the effort, Its own quality indicators are compatible with those of the NCQA Initiative Indicators planned for the standardized report card include childhood vaccination rates, breast and cervical cancer screening rates, and hospitalization rates for pediatric asthma cases (941). A preliminary version of the report card IS projected for completion by the end of 1994 (48)

Second, the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) announced, in 1993, the introduction of the first two sets of quality indicators for a program in which hospitals could participate on a voluntary basis; participation in the Indicator Monitoring System IS to become compulsory in 1996 (429) The data collected by JCAHO from Individual hospitals wiII be translated into scores (based on compliance with recognized standards of care, such as clinical guidelines) in 50 areas. The scores will be available for use by the hospitals themselves as well as by consumers (553)

It is Important to note that, while most private sector quality initiatives have focused on producing report cards that may be used by employers, the legislative language of several health reform proposals implies that the explicit audience for quality assessments should shift to the individual, who will be choosing coverage from a selection of plans made available through a purchasing cooperative (S 1757, H.R 3222, S 1770),

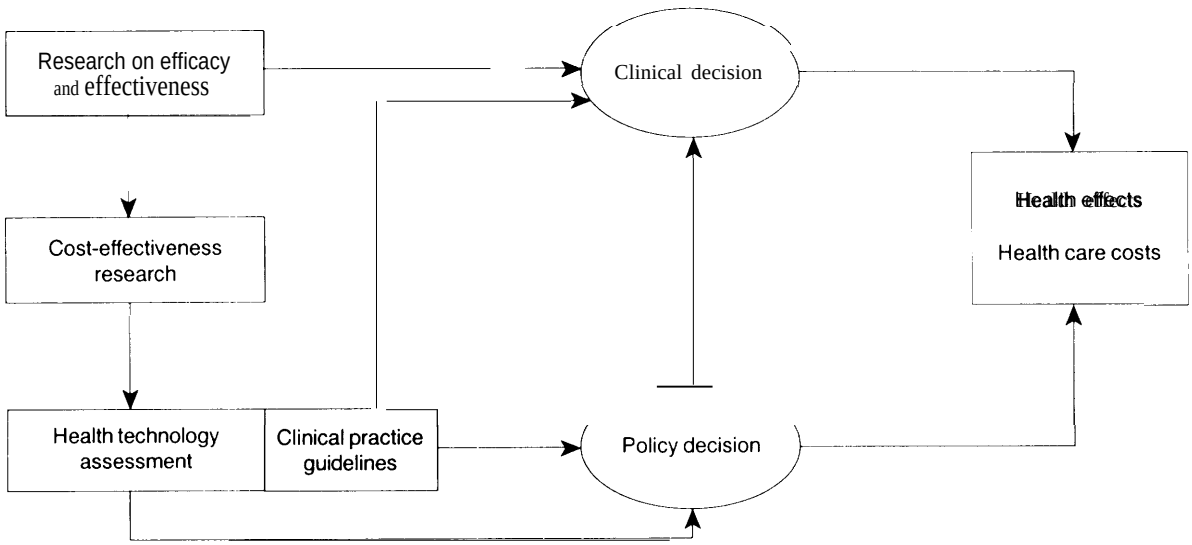
The switch from the employer to the Individual as the unit that generates and controls demand for health care coverage raises additional issues to confront in the development of report cards Whether the level of interest in using report cards on the part of individual consumers will be comparable to that of employers is still unclear (553) Some observers have argued that outcomes-based report cards wiII not be easily interpreted by consumers unfamiliar with medical issues (566) The extent to which other factors, such as personal relationships with physicians and the recommendations of peers, may compete with or outweigh the value of report cards in individual decision-making is also unknown (553)

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

of effectiveness research, assimilating existing research and adding educated opinions where research results are lacking or controversial. And a major purpose of most guidelines is to inform an individual clinical decision directly. At the same time, guidelines also often provide a basis for broad clinical policies (e.g., in a medical practice) and insurance coverage policies, and they certainly can include the consideration and analysis of information other than effectiveness information (e.g., costs).

In this report, clinical practice guidelines that address medical technologies and practices, and that are created through a structured format of synthesis and analysis, are considered a special and particularly relevant category of health technology assessments. Many of the clinical guidelines discussed here are undertaken in order to guide the formation of a clinical policy rather than a purchase decision or insurance coverage policy, and information on effectiveness is often (he predominant concern. but this report

**FIGURE 2-1: A Framework for Evaluating Health Care Technologies**




SOURCE Office of Technology Assessment 1994

does not consider such efforts to be part of effectiveness research itself.

## THE SHAPING OF EFFECTIVENESS RESEARCH

The prime contributor to the current enthusiasm for effectiveness research, and for the use of particular tools and methods in that research, derives from fertile field of health services research. This field first became a recognized discipline in the late 1960s, as it brought together people from diverse social science and clinical backgrounds with interests in untangling the underlying factors affecting the patterns, quality, and cost of health care.

Much of this research comprised studies that investigated relationships within the health system as a whole. Relationships between peoples' access to care and health status, between trends in health care services and trends in health care costs, and other subjects relating to the cost, quality, and accessibility of care are longstanding areas of health services research. One segment of research into the patterns and quality of care, however, developed lines of inquiry that began to focus on the patient-level consequences of clinical care. This line of research, which received its impetus from intriguing findings about variations in clinical practice across geographic areas, led to a number of different research efforts examining the appropriateness and outcomes of patient care, and it ultimately resulted in the federal government's medical effectiveness research initiative.

### Geographic Variation in Medical Practice

Research into geographic patterns of care was one of the earliest areas of health services research. Variations in the rate at which patients use medical services, and the rate at which physicians perform them, have been an intriguing topic of health care research for decades. A seminal study by Glover in the 1930s showed that the percentage of British schoolchildren who had undergone tonsillectomies varied more than tenfold across areas of England and Wales (285).

Studies of surgical procedures in the United States and Canada in the 1970s and early 1980s documented similarly large differences across small geographic areas. Hysterectomy rates var-

## BOX 2-4: Variation Between Two Cities: The Case of New Haven and Boston

In a frequently cited 1987 article, Wennberg and colleagues [Illuminated a dramatic example of geographic variation in the utilization of Inpatient care (91 7) New Haven, Connecticut and Boston, Massachusetts are demographically similar cities in which most hospital care IS provided by academic health centers However, 1982 per capita expenditures for inpatient care in Boston were roughly twice those of New Haven ($451 vs $889), making the Boston community one of the biggest consumers of health care services and New Haven one of the smallest in New England The authors estimated that about 80 percent of the increased utilization in Boston was attributable to higher hospital admission rates (as opposed to greater lengths of stay) A look at rate differences for specific procedures and operations found ratios of utilization to favor New Haven in some cases and Boston in others

These observations led the authors to ask whether hospital services were being rationed in New Haven or over-utilized in Boston An assessment of hospital resources in the two cities showed that Boston residents were allocated 55 percent more beds per capita than residents of New Haven The researchers wondered whether New Haven suffered from a shortage of beds, which would force doctors to consciously control rates of hospital admission However, based on conversations with physicians, as well as the observation that New Haven hospital bed occupancy rates averaged only 85 percent, Wennberg and colleagues concluded that this was not the answer

Wennberg has suggested that Boston's higher hospital admission rates, and thus its higher health care expenditures, might be the result of a need to cover the cost of maintaining additional resources in the form of beds, personnel, and equipment (91 2). In 1989, 38 hospital beds were available per 1,000 Bostonians, the statistic for New Haven residents was 26 per 1,000. Because there IS no evidence that the additional expenditures are linked with better outcomes for patients, the Boston-New Haven example has been evoked often to support the view that a significant fraction of health care spending may be wasted on unnecessary care.

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

ied five-fold across hospital service areas within the province of Manitoba, Canada, for example (646). Tonsillectomy rates in Vermont varied from three to 15 tonsillectomies per 1,000 residents (91 4). Similar variations across small areas in other states were found for procedures such as hernia repairs, appendectomies. and Cesarean sections (39,456,9 15). Considerable variations occurred across states as well. Rhode Island physicians, for example, performed twice as many hysterectomies and prostatectomies per thousand population as did physicians in Maine (39).

Variations were not limited to surgical procedures. Rates of hospitalization for back injury varied tenfold across Maine (909), and physicians in Boston hospitalized adult patients with medical problems at one and a half times the rate of physicians in New Haven (917) (box 2-4). Rates of use of both medical and surgical procedures by Medicare beneficiaries varied tremendously both within and across states (11 7).

Along with the accumulating evidence of great variations in the use of medical services was an increasing awareness through the 1980s that these variations were not necessarily associated with discernible differences in the need for health care among different populations. Areas showed great differences in practice variation even when they

had populations that were similar in both their demographics and their measurable rates of morbidity, such as days in bed due to disability (649, 908,909,917).

## Outcomes of Patient Care

A critical question raised by the research on practice variations was whether these differences in medical practice were associated with corresponding differences inpatient outcomes (916). In one of the earliest studies examining this question, Daniels and Schroeder found no relationship between physicians frequency of laboratory test use and the degree to which their hypertensive patients' blood pressures were under control (149). Other studies suggested that for some surgical services, having surgery was not associated with a decreased risk of death (646,915).

Where differences in medical practice could not be linked to differences in underlying health needs or to differences in health outcomes, researchers theorized that the rate of procedures in the high-rate areas could be lowered, and costs reduced, while maintaining good patient outcomes. Conversely, where different rates were associated with different outcomes, then overall patient outcomes and the quality of care could theoretically be improved by moving practice towards the best-outcome rate. In either case, physician preferences and uncertainty appeared to be a major determinant of the procedure rate in any given community (196,9 11). If these factors were indeed at the root of practice variation, then the tantalizing possibility arose that many instances of medical intervention might be avoided, and better health outcomes achieved, simply by more agreement on the best course of care.

### *The Study of Outcomes of Prostate Disease*

In the early 1980s, Wennberg and colleagues began to focus specifically on the study of the treatment of benign prostatic hyperplasia (BPH), a noncancerous enlargement of the prostate gland that is a very common condition in older men. Prostatectomy is one treatment for BPH, and the proportion of men undergoing the procedure by age 85 varied from as low as 10 percent to as high as 50 percent in neighboring communities (911).

Discussions with physicians confirmed the existence of two differing views within the physician community. One view held that prostatectomy should be performed as early as possible after diagnosis. Doing so, these physicians maintained. would avoid the development of later symptoms, and the need for surgery when the patient was older and the procedure riskier. The opposing view held that surgery did not improve overall life expectancy and should be reserved for patients with symptoms. When the researchers examined the literature and insurance claims data on treatments for BPH and compared them with mortality rates, they concluded that surgery did not improve life expectancy and might actually decrease it (911 ).

Besides creating an interdisciplinary approach to research that focused on patient outcomes. "outcomes research" as defined by the activities of the BPH project had three characteristics that shaped the directions of subsequent effectiveness research.

First, the project drew attention to the differences between outcomes predicted by clinicians, for various alternative therapies, based on their knowledge and experience, and the outcomes experienced by patients, as documented in the data. Neither existing literature nor expert opinion on prostatectomy suggested as high a mortality rate as that found for the patients represented in the claims database used by the BPH research team (918).

Second, it made great use of insurance claims data as a basis for assessing the actual outcomes associated with particular therapies in practice. The identification of greatly varying rates of prostatectomy through insurance claims data was responsible, in part, for the decision to focus on BPH as a condition to study (248). The BPH study also found higher reoperation and mortality rates associated with an increasingly popular, less invasive form of prostatectomy (transurethral prostatectomy, or TURP), based on an analysis of claims data for the procedure (920). Those results gained wide publicity as an unexpected finding that

demonstrated the benefits of this observational data-based approach to documenting outcomes. Although this association was later shown to be due at least in part to patient selection bias (physicians tended to refer higher risk patients for the less invasive surgery and reserve open surgery for lower risk patients) (248), the initial finding nonetheless helped promote the use of claims data as a method for studying patient outcomes.

Third, the BPH outcomes project placed a heavy emphasis on understanding patient preferences and measuring patients' self-reported symptoms and quality of life. Given that therapy for BPH is targeted at reducing or preventing symptoms rather than improving lifespan, and given the lack of clear objective benefit, based on existing studies, of one therapy over another, BPH researchers concluded that patient preferences should be a major component of the decision to select a particular mode of therapy (911).

### *Relationships Between Volume and Outcomes of Care*

A separate cadre of researchers homed in on another aspect of variation in medical practice: the relationship between the volume of a procedure done in a hospital. or by an individual physician. and the outcomes of the patients who underwent that procedure. The common theme of this body of literature is that there is often a correlation between the number of procedures performed by a provider and the outcomes of care (generally measured by mortality rates).

Luft and colleagues published a landmark study in 1979, suggesting that, at least for some procedures, higher hospital surgical volumes were associated with better outcomes for patients (476). They compared mortality rates with surgical volume of 12 procedures for nearly 1,500 hospitals during 1974 and 1975 and found that for open-heart surgery, vascular surgery, transurethral resection of the prostate. and coronary artery bypass graft (CABG) surgery, high-volume hospitals (defined as hospitals that performed a given procedure 200 or more times annually) had mortality rates 25 to 41 percent below their low-volume counterparts. For four other procedures, researchers found that the volume-outcome curve flattened out at a much lower annual volume threshold (10 to 50 procedures per year as opposed to 200). In two cases. no volume-outcome relationship was observed.

Subsequent studies confirmed the finding that hospitals with more experience in a procedure-i.e., higher volumes of surgery—had significantly lower rates of in-hospital mortality (256.257, 426,693 ). Few equivalent volume-outcomes studies on medical conditions have been performed, although two studies of AIDS treatments found that patients with AIDS fared substantially better at hospitals serving large numbers of patients with AIDS cases, compared with their counterparts at low-volume hospitals (47,732).

Not all studies investigating possible volume-outcome relationships have found them. A 1987 review of the literature regarding this relationship for hospitals found that unlike most studies of other procedures, studies of treatment for femur fracture and for stomach operations tended not to support the "greater volume-better outcomes" hypothesis (477).

The research on volume-outcomes relationships emphasized the usefulness in health research of ultimate measures of health outcomes, such as mortality, rather than intermediate endpoints with less clear functional implications. Although the exact nature of the relationship between volume and outcomes remains murky. the research overall has tended to reinforce the idea that simultaneously reducing costs (through improved efficiency at high-volume institutions) while improving the quality of care (through better care outcomes) is an achievable goal.

### *The Medical Outcomes Study*

The theme of improving measurement of patient outcomes gained substantial support from an entirely separate and ambitious research initiative. the Medical Outcomes Study (MOS), which began in 1986. The goal of the MOS was to follow the health care received by a large group of participants in order to answer outstanding questions

about the relationships between the structure and process of care and the health outcomes associated with that care (746). To do so, the MOS researchers collected cross-sectional (i.e., one-time) data on over 22,000 participants. In addition, the researchers identified a subset of over 2,000 patients who had at least one of five conditions (hypertension, diabetes, acute myocardial infarction, congestive heart failure, and depression) and began collecting detailed longitudinal data on their care. Data collection on these patients was still ongoing as of 1993 (745,746).

To assess the outcomes of care on patients, researchers used information from clinical examinations and from the patients' medical records (746). In addition, the researchers developed and tested at length a set of general health surveys, administered to patients, to assess the patients' own perceptions of their functioning and general well-being.[4]

The MOS made two crucial contributions that helped give focus to effectiveness research efforts. The first was its substantial investment in developing and validating general health measurement instruments, particularly the 36-question version, the "SF-36," to measure self-assessed patient functioning and wellbeing at any point in time (513). The second contribution was to link patient characteristics and particular components of care with care outcomes (725). Researchers have found, for example, that the negative effects of depression are additive for patients who are depressed in addition to having other chronic health problems (907).

## Appropriateness of Care

Even when an intervention is generally effective, or effective under particular circumstances, it may sometimes be applied to patients for whom it is inappropriate. The research on variations in medical practice led directly to another question: Does the greater inappropriate use of procedures in high-use areas explain the geographic differences in rates of use?

There has long been evidence that some inappropriate medical practice does occur (242). A very convincing study done in the 1970s, for example, documented the inappropriate use of tetracycline, an antibiotic, among young children in Tennessee's Medicaid program (626). Complications related to the use of tetracycline in this age group had long been noted, and by the 1970s there were numerous alternative drugs. In January 1975, the American Academy of Pediatrics officially stated that there were "few if any reasons for using tetracycline drugs in children less than 8 years old." Despite the uniform agreement in the official medical community regarding tetracycline's inappropriateness for children in this age group, Ray and colleagues found that the drug had been prescribed for over 4,000 young children over a two-year period (626).

Researchers at RAND approached the question of appropriateness of care by focusing on specific procedures that are both costly and shown to vary across geographic areas (11 8). Initially, they chose six procedures to study:

1. coronary angiography (a diagnostic imaging procedure for heart disease),
2. coronary artery bypass graft surgery (a major surgical treatment for heart disease),
3. cholecystectomy (surgical treatment for gallstones),
4. diagnostic gastrointestinal endoscopy (a procedure to diagnose disorders of the digestive tract),
5. colonoscopy (a diagnostic procedure to detect disorders of the lower intestine), and
6. carotid endarterectomy (a surgical procedure performed in persons considered to be at very high risk of stroke).

A major obstacle to overcome was defining "appropriate" uses of these procedures. Unlike the tetracycline study, which had the advantage of an

[4] In an interesting example of the accumulative properties of health services research, the foundation for the health surveys was an assessment measure from a previous major federally funded research effort, the RAND Health Insurance Experiment (86).

## BOX 2-5: Defining "Appropriate"

To define "appropriate'" indications for the procedures they studies, RAND researchers convened expert panels that reviewed the indications discussed in the literature, and in their own experiences, and arrived at group ratings of the appropriateness of each indication (see chapter 7) The panels used a rating scale of 1 through 9, with 1 representing extremely inappropriate and 9 representing extremely appropriate "Appropriate" was defined to mean that the expected health benefit exceeded the expected negative consequences "by a sufficiently large margin that the procedure was worth doing " "Inappropriate" meant that the negative consequences outweighed the health benefits Panelists were instructed not to consider financial costs

The researchers suggested a final, simpler spilt to categorize the ratings three categories of Inappropriate, " "appropriate, " and "equivocal " The definition of the last category was particularly Interesting, because it included both indications for which the panel agreed that the Indication was neither clearly appropriate nor clearly Inappropriate, and Indications for which there was substantial disagreement among panelists regarding appropriateness

SOURCE Off Ice of Technology Assessment 1994 See chapter 7 and appendix C text for more detailed discussion and reference sources

unambiguous measure of appropriateness in the statement of a major medical association, there was no universally acknowledged consensus about what constituted appropriate use. The issue was not that these procedures (e. g., bypass surgery) were themselves inappropriate, but that some of the reasons for doing them—the medical indications—were not appropriate. To define '*appropriate" reasons for performing the six procedures, researchers at RAND assembled "expert panels'* to rate the various identified medical indications for each procedure (box 2-5).

The results of applying appropriateness ratings to explain geographic variations in medical practice have been somewhat surprising. In the first study on this topic. researchers examined the reasons for performing three of the six procedures (carotid arterectomy, coronary angiography, and gastrointestinal endoscopy) in five sites across the country (11 8). The rates at which each of the three procedures were performed varied considerably across sites (in the case of carotid endarterectomy, they varied by almost a factor of four). The proportion of procedures performed "inappropriately" according to RAND criteria, however, was surprisingly consistent across sites (between 29 and 40 percent for carotid endarterectomy and between 15 and 19 percent for the other two procedures). Overall, there was an association between higher rates of use of a procedure and a higher proportion of inappropriately performed procedures, but the association was surprisingly small (1 18).

To test the possibility y that the use of large areas for comparison might have masked variations that would be apparent if smaller areas were contrasted, the researchers repeated the process in 23 counties in a single state (447). Both the rates of procedures and the percentage of procedures rated appropriate varied enormously across these small areas (table 2-1 ). Nonetheless, the association between the two measures was remarkable for its near absence (447). Although this study has been criticized as inadequate to test its hypothesis properly (152), its findings were so remarkable that they are hard to dismiss out of hand.

Using the RAND appropriateness criteria, the same group of researchers have documented significant proportions of inappropriately performed procedures in several patient populations (525a, 938,939). In a literature review that included these and other investigations into inappropriate care.

**TABLE 2-1: Variation and Appropriateness of Three Procedures Across Small Areas**

| Procedures | Rate of use per 10,000 Medicare enrollees | Percent of procedures judged appropriate |
|---|---|---|
| Coronary angiography | 13-158 | 8% - 75% |
| Carotid endarterectomy | 5-41 | O% - 67% |
| Upper gastrointestinal tract endoscopy | 42-164 | 07. - 25% |

SOURCE Based on data from L L Leape, R E Park, D H Solomon, et al "Does Inappropriate Use Explain Small-Area Variations in the Use of Health Care Services')" *Journal of the American Medical Association 263(5) 669-672, 1990*

the reviewers found documentation of inappropriate use ranging from 3 to 75 percent for procedures, 6 to 80 percent for hospital use and office visits, and 3 to 90 percent for drug use (83). They found evidence of underuse as well as overuse, although the latter was much more prevalent in the literature. They concluded by speculating that:

> ... as much as one-fifth to one-quarter of acute hospital services or procedures were felt to be used for equivocal or inappropriate reasons, and two-fifths to one-half of the medications studied were overused in outpatients (83).

International comparisons suggest that even countries with much lower overall rates of procedures than the United States have a substantial proportion of procedures that are performed for inappropriate reasons. Physicians in the United Kingdom, for example, perform coronary angiography and coronary bypass surgery much less frequently than do U.S. physicians (54). As expected, in a study comparing the appropriateness of indications for these two procedures in the two countries, researchers found that U.K. physicians rated a higher proportion of indications to be inappropriate for both procedures than did U.S. physicians (54,85) Nonetheless, the proportion of these procedures deemed inappropriate even by the U.S. physician panel was a substantial 17 percent (54).

Overall, the findings of appropriateness research have tended to support the belief that some portions of medical care can be eliminated while actually improving the quality and effectiveness of care provided. That belief may be somewhat overstated. The main message from the RAND review of appropriateness studies—that up to one-fourth of procedures and up to one-half of medications are prescribed for reasons that are inappropriate or equivocal—may imply more "wasted" care than is the case. The selection of technologies that have been studied, for example, may be biased if researchers have tended to study a particular technology or service precisely because inappropriate use was suspected.[5] In addition, the reviewers' generalization of appropriate use combined equivocal with inappropriate care. Recent studies suggest that the category of equivocal care is sometimes much larger than the category of care that is clearly inappropriate (54,343,446).

Nonetheless, appropriateness research has certainly documented that a significant amount of dubiously useful care is being provided. This research has also helped highlight the degree of professional uncertainty and disagreement that remains in the appropriate indications for performing many high-cost procedures. But the findings of this research also suggest that areas with high

[5] Indeed, the one study reviewed by the RAND researchers that looked at a broader set of 12 procedures found a much lower rate of overuse (3 percent) than any of the studies looking at overuse of a single procedure (the lowest rate of inappropriateness found in any of these studies was 13 percent).

rates of particular procedures do not necessarily have a higher proportion of inappropriate procedures, In fact, areas with low rates may not perform et-tough appropriate procedures.

The research addressing the question of the appropriate use of particular medical technologies has diverged somewhat from the line of research that makes up most of the federal medical effectiveness program. Unlike the work on patient care outcomes, the extensive RAND work on appropriateness of care has focused more on the pragmatic demand for information that can lead to immediate, relatively unambiguous decisions.

The ability to label some care as "inappropriate" is potentially useful to policy makers interested in taking immediate action to reduce some proportion of health care costs through the elimination of "wasteful" services. The attractiveness of the RAND approach is apparent in the fact that private sector payers and providers are expressing an interest in linking medical practice guidelines and payment to conclusions about appropriateness based on this approach. The limited assessment of this approach outside of the small group of researchers who developed it, however, has led some observers to criticize the adequacy of its evaluation (605). (Chapter 7 and appendix C of this report discuss the process used in the RAND approach in more detail. )

## The Federal Medical Treatment Effectiveness Program

The different lines of inquiry into the variation and outcomes of current medical care practices began coalescing into a program identity in the late 1980s. Encouraged by the progress of research into the outcomes of treatments for prostate disease. Congress in 1987 ordered the National Center for Health Services Research (NCHSR)[6] to establish an "outcomes research program" to expand this approach to understanding medical care. NCHSR solicited applications for the first outcome research team program grants in 1988.

In the same year, William Roper, then administrator of the Health Care Financing Administration (HCFA), and several of his colleagues issued a call for "effectiveness research" (651 ). Roper's focus was on the effectiveness of medical care provided to elderly and disabled individuals covered by Medicare. He proposed to examine the outcomes of medical procedures and other care by making use of the rich resources that were the Medicare administrative databases (65 1).

The creation of the Agency for Health Care Policy and Research (AHCPR) by congressional mandate in 1989 eclipsed Roper's plans for a HCFA research initiative. A major part of AHCPR's role was to be the focal point for federally supported effectiveness research. To carry out this role, AHCPR established its Medical Treatment Effectiveness Program (MEDTEP), which subsumed both the HCFA initiative and the previous NCHSR outcomes research program. Research into practice variation and documenting outcomes of current medical practice continued to be part of the research portfolio.

Although "effectiveness research," as defined earlier in this chapter, could cover a very diverse set of research activities, the characteristics of the federal government effectiveness initiative have been shaped by its roots in research on practice variation and the measurement of health outcomes. (Research on the appropriateness of care, as carried out at RAND, has been carried out separately from the federal initiative.) The outstanding characteristics of the federal endeavor based at AHCPR are:

1. It is focused primarily on the evaluation of existing technologies and medical practice patterns, rather than on the evaluation of new interventions.

[6] NCHSR went through several name changes between its inception in 1968 and its replacement by the Agency for Health Care Policy and Research in 1989. In 1987 its formal name was the National Center for Health Services Research and Health Care Technology Assessment, but for the sake of simplicity the shorter title is used here and elsewhere in this report.

2. It has emphasized the need for research whose results will be widely applicable, including populations and settings that have often been underrepresented in efficacy studies. These include elderly populations, women, minorities, persons with disabilities or multiple health problems, and treatment settings such as physicians and health facilities that are not affiliated with teaching institutions.
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 only intermediate clinical measures (e.g., laboratory test scores).
4. It has included the substantial use of tools other than prospective, randomized clinical trials. In particular, it has historically placed a particular emphasis on 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 has come from the expectation that for existing medical technologies, nonclinical research methods are both cheaper and more efficient.

## EXPECTATIONS IN THE CONTEXT OF NATIONAL HEALTH REFORM

Effectiveness research stresses that medical practice varies for reasons unassociated with demographics and health needs, and that much current medical care is performed for inappropriate or at least equivocal reasons. If this is true, and if the most effective practices can be identified, described, and disseminated, then it might indeed be possible to raise the quality of health care while reducing its costs. This is the basic assumption that underlies many of the expectations of effectiveness research. It is also the assumption that led the federal government to invest substantially in the creation and dissemination of clinical practice guidelines, which would assemble the evidence and describe the best course of clinical care for the medical conditions they addressed.

The assumption found a ready audience in public policy makers, embroiled in the search for palatable solutions to the conundrum that is American health care. Since the early 1980s, researchers and commentators have promoted the idea that pursuing research into the effectiveness, cost-effectiveness, and broader effects of health care would be a small investment yielding a major improvement in both the quality and the cost of care (8,791a,908a,934). The message was clearly heard by members of Congress. At a Senate hearing in 1988, the opening statements of the Senators reflected a confidence that health services researchers would be able to define appropriate care in order to offer substantial cost savings and high quality, focusing on the advantages for Medicare beneficiaries (792). It was against the background of these expectations that AHCPR was created in 1989 to provide focused federal support for effectiveness research and clinical practice guideline development.

Since the establishment of AHCPR, the rhetoric emphasizing the cost-containment benefits of these activities has faded somewhat. Cautious notes have been sounded by reports from the Institute of Medicine and the Physician Payment Review Commission, which backed the idea of federally supported guidelines but questioned whether AHCPR's guidelines effort would necessarily lead to cost savings (376,607). Recently, the administrator of AHCPR has asserted bluntly that "outcomes research is not a cost cutting exercise" (494).

Nonetheless, with the prospect of national health reform on the horizon, effectiveness research, guidelines development, and other activities that involve the evaluation of clinical practices continue to play a part in policy makers' hopes for improving the health care system. President Clinton's health proposal, for example, included specific provisions to encourage "effectiveness research," "quality and outcomes research," and the "development and dissemination of guidelines." According to the proposal, this research would "increase the cost-effectiveness, appropriateness and quality of care" in the health care system (S. 1757).

Of particularly widespread interest in health reform proposals is the idea of "scorecards" or "per-

**BOX 2-6: Examples of State Legislative Activities Regarding Clinical Practice Guidelines**

- **Maine** has an ongoing five-year demonstration project that permits the use of guidelines as a defense in malpractice cases (453)
- **Vermont** has a similar law that calls for "recommendations for the development of standards of care and practice guidelines, ' which could be used as a defense in malpractice suits (Vermont Law Sec 1, 18 V.S.A Part 9, Ch 221)
- **Maryland** has established an Advisory Committee on Practice Parameters to oversee the design of guidelines whose content are to be based on effectiveness research and physician consensus
- In **Minnesota,** the Health Right Act of 1992 included the adoption of practice parameters as a means of assuring quality in health care Here, too, guidelines may be used as a defense in malpractice suits, and the fiscal expectations for guidelines are eloquently demonstrated by the fact that the Minnesota Department of Health listed the provision for practice parameters under the heading of cost containment, as a measure "to avoid unnecessary and Ineffective treatment and services (533)

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations at the end of this report

formance indicators," which rate providers, insurers, or health care plans according to their performance along several criteria. These can include mortality rates, costs, rates of specific procedures, or rates of hospitalization for "preventable" diseases. A number of payers and providers already have systems in place to monitor performance indicators (see box 2-3). They see these systems as strategies to eliminate costly or inadequate physicians from the payment rolls, improve hospital quality of care over time, and assure premium-paying employers that their health care dollars arc being well spent. In this case. the results of effectiveness research and clinical practice guidelines can become the benchmarks against which providers are rated.

Clinical practice guidelines have also become a basis for policy makers' hopes of reducing malpractice insurance costs and physicians' use of defensive medicine (e.g.. H.R. 101), especially at the state level (box 2-6).[7] Perhaps most importantly, guidelines and data on effectiveness have also been proposed as the basis for defining health insurance benefits.

In perhaps the best known example, the State of Oregon, in 1989, officially proposed prioritizing health care services for its Medicaid beneficiaries according to such factors as the relative effectiveness of the services (722). Although in the end evidence on effectiveness played a relatively minor role in the prioritization process (788,794). the process shaped the discussion about the place of information on effectiveness, cost-effectiveness. and quality of life in health insurance coverage. More recently, legislation introduced in Oregon would require that medical guidelines be part of the basis for prioritizing services under the state's Medicaid demonstration program (Oregon Senate Bill 757. 1993).

[7] "Defensive medicine" occurs when doctors order tests, procedures or visits, or avoid high-risk patients or procedures, primarily (but not necessarily solely) to reduce their exposure to malpractice liability (790). For a detailed discussion of this topic, see the OTA report, *Defensive Medicine and Medical Malpractice* (790).

## BOX 2-7: One Proposed Model for Basing Health Insurance Benefits on Clinical Practice Guidelines

In one model of how a benefits package might be based on clinical practice guidelines, Hadorn has proposed the development of a comprehensive set of "necessary care guidelines, " which would collectively represent a basic benefits package (318,320), He defines "necessary" as "reasonably well demonstrated to provide significant health benefits, " one step beyond appropriateness (320,403). Under this model, necessary care guidelines, resembling utilization review criteria in format, would be developed by expert panels and presented for public debate at hearings modeled after the '(science court experience" and the NIH Consensus Development Conferences (31 8,319)

Hadorn's proposal hinges on the ability to incorporate into the benefits development process an "objective standard of proof" that would consider health care needs as well as costs, thereby constructing a mechanism to judge a given type of care on the "net health benefits" that the population could expect to gain from it (319), Gwen that the goal of this model is to provide comprehensive coverage while cutting costs, the major assumption is that it would be unnecessary to make decisions based on cost alone (i e rationing) because "the volume of services excluded from coverage using a standard of proof approach would entail substantial cost reduction in and of Itself"' (319)

SOURCE Off Ice of Technology Assessment 1994, based on sources as shown Full citations are at the end of the report

Other public and private policy makers have also begun to experiment with the use of clinical practice guidelines in defining or modifying health insurance benefits. In Canada, a preliminary agreement with the British Columbia Medical Association stipulates that patients who seek services outside the parameters of practice guidelines (now being developed by the province's Medical Services Commission) will not be covered by Canada's national health insurance (528). In the United States, Blue Cross and Blue Shield of Illinois implemented, on January 1, 1994, a policy requiring physicians to comply with practice guidelines. Participating specialists in the Illinois Blues' Managed Care Network Preferred, servicing over 100,000 enrollees, must follow guidelines covering 14 procedures or treatments, including bypass surgery, cholecystectomies, and blood transfusions. Except for guidelines on cancer care, which were developed by the insurance company, the practice parameters were produced by various specialty societies. The new policy was met with opposition from the American Medical Association, which argued this mechanism made physicians who participated in the medical plan subject to guidelines that they had no opportunity to help develop or modify (74).

The Clinton Administration's proposal for national health care reform also incorporated effectiveness and cost-effectiveness research results into its proposed benefits plan, at least for preventive services (S. 1757). Both this and alternative proposals that involve the establishment of a national board that would set benefits clearly envision that such a board would use the results of effectiveness and cost-effectiveness research and of clinical practice guidelines and other technology assessments in their decisionmaking (e.g., S. 1757, S. 1579, H.R. 3222). Some researchers have taken the concept a step further and proposed an insurance benefits model in which a battery of guidelines would themselves comprise a benefits package (box 2-7). In California, policy makers have considered using guidelines to create a benefits package for the state insurance plan for public employees (99).

If data on effectiveness and formally structured clinical practice guidelines are one of the bases for health insurance benefits under health reform,

then the validity and reliability of those inputs are clearly of considerable interest. Even in the absence of a benefits package that relied heavily on research-based evidence and guidelines, any reform proposal that relies on the expansion of "managed care" has a stake in the validity and impact of these activities. They represent some of the tools by which the managers in managed care organizations can hope to achieve high-quality, better-cost care. If these tools are inadequate, the assumption that managed care can solve many of America's health care problems would bear serious scrutiny.

## CONCLUSIONS

Much, if not most, of existing medical technology and practice has been inadequately evaluated, even with regard to its effectiveness in improving peoples' health. Nonetheless, for all this dearth of information, society has gradually amassed a number of tools to evaluate the health, economic, and social effects of technologies (366,783), and the applications of those tools to the crucial questions of health care are slowly growing.

Research to address the deficit in evidence regarding current medical care has developed separately from the traditional clinical trials research community, influencing the kinds of tools it has applied. The research evaluating existing clinical practices has also tended to emphasize that considerable variations exist in how medical care is practiced; that considerable disagreement exists among clinicians regarding the circumstances under which particular treatments are appropriate; and that the health outcomes valued by patients are often not the same as those measured by researchers and clinicians. In the process, effectiveness research has created expectations among policy makers that further investments in this line of research. coupled with the aggressive development and promotion of clinical practice guidelines, can make great strides in eliminating ineffective care, improving the overall health of the population, and even reducing health care costs.

Despite the optimism prompted by early effectiveness research, there were and are still a number of ambiguities about the kind of change that can be expected. The research on appropriateness, for example, has found that higher rates of use of a procedure are not equal to higher levels of inappropriate care. Nor does current research necessarily support the idea that the source of variations in clinical practice is individual provider uncertainty that can be abolished by presenting that practitioner with good information or guidelines. Rather, research suggests that uncertainty lies in disagreements among physicians (459), with individual physicians possibly quite confident in their own opinions. Indeed, Chassin (11 5) theorizes that the main reason behind practice variation is the number of physicians who are "enthusiasts" for particular procedures or care processes. If this is true, there may be disagreement but not individual uncertainty, implying a more difficult job for federally sponsored activities whose ultimate goal is to affect clinical practice by improving outcomes, reducing costs, or both.

Implicit assumptions about the impact that these activities will have underlie a number of different aspects of proposals currently being discussed in the context of national health reform. It will affect, for example, the extent to which policymakers can depend on the idea that basing health benefits on guidelines and effectiveness information is feasible. likely to result in changes in clinical practice, and likely to help restrain system costs.

Moreover, the findings of effectiveness research and practice guidelines are a crucial underpinning of performance indicators, which are based on the idea that there is a proven standard of preferred practice to which a provider should adhere. Health reform proposals that emphasize a large role for other consumer and provider information, or for managed care providers, contain implicit assumptions that evidence regarding the

effectiveness and value of medical technologies and practices is sufficiently available, valid, and convincing that it will enable these players to improve their health care outcomes and costs.

As implemented in the federal government's medical effectiveness initiative in 1989, and in the charge to AHCPR, "effectiveness research" emphasized particular qualities and approaches toresearch. Those qualities (e.g., an emphasis on existing technologies and broad populations) and approaches (e.g., large database analyses) were emphasized in response to perceived deficits in the contemporary research agenda. However, "effectiveness research" includes a wider variety of potential activities than those emphasized in the first few years of AHCPR's existence.

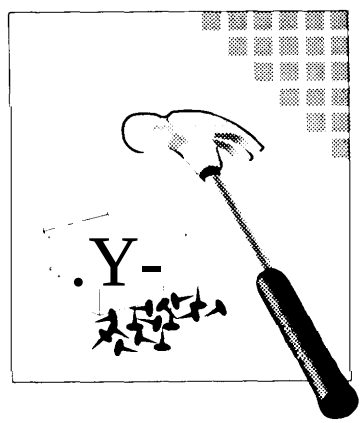
Examining the federal government's current investment in activities that evaluate the effectiveness and value of medical technologies and practices in detail, and examining the extent to which expectations for that investment are well founded, is the focus of this report. The remainder of this report assesses the validity, potential usefulness, and efficiency of Federal activities regarding effectiveness research (chapters 3 and 4), cost-effectiveness research (chapter 5), health technology assessment generally (chapter 6), and clinical practice guidelines specifically (chapter 7). Finally (chapter 8), this report examines the ways in which these activities, and particularly clinical practice guidelines, are most likely to have an impact on clinical practice.

# 3 Tools for Effectiveness Research

The strength and believability of evidence on the effectiveness of health technologies rest largely on the underlying methods used to generate it. The purpose of this chapter is to describe the basic methods employed to generate evidence, emphasizing those techniques that have evolved recently, are particularly appropriate to broad research on the effectiveness of care, or have seen especially heavy use in effectiveness research as carried out thus far in the United States. [1]

The validity of the underlying methods being applied is a matter for particular interest in the area of research on the effects of health technologies, because making heavier use of certain techniques was an explicit component of the federal government effectiveness initiative. The law enacting the Agency for Health Care Policy and Research (AHCPR) specifically encouraged the use of particular research methods, such as large administrative database analysis (Public Law 101 -239). In addition, the increasingly intense interest in whether specific medical interventions are worth doing has stimulated research activity in areas ranging from the measurement of people's preferences for various health outcomes to the statistical synthesis of the results of pre-existing studies.

Legislation to encourage effectiveness research has not only encouraged certain research methods but also a particular organizational structure for applying them. This approach centers on Patient Outcomes Research Teams (PORTs)-groups of research-



[1] More detailed discussions of the applications, advantages, and limitations of some of these techniques are contained in a separately published set of background papers associated with this report (see appendix A).

ers with diverse backgrounds who join together to conduct research on a particular medical condition. This chapter concludes by examining the contributions of these teams and their implications for the future of effectiveness research.

## TECHNIQUES TO EVALUATE HEALTH EFFECTS

Tools for generating evidence regarding the effectiveness of health technologies fall into three broad categories:

- *basic tools* for measuring health status and health outcomes;
- *primary studies,* such as clinical trials and administrative database analyses; and
- *secondary techniques* to synthesize the results of the primary studies in order to generate new insights or more powerful conclusions.

### Basic Tools for Measuring Health Outcomes

Assessing the effectiveness of a medical technology (or any health care intervention) requires an evaluation of whether the health-related outcomes resulting from the use of that intervention are better than would have been expected without it. Such an evaluation requires measuring both what those outcomes are and what they would have been without the technology.

The simplest outcome to measure is death. For some conditions, it is probably the most important outcome as well. But most of the conditions that cause people to seek medical care, such as back pain and bronchitis (795), are not characterized by high fatality rates. Furthermore, even for conditions that are often fatal (e.g., cancer), improvements in the quality of life people have before death is a major goal of treatment (820).

An interest in measuring people's health status more directly has led to the development of tools to assess how patients feel and how well they can function. At the most basic level, existing tools differ according to two attributes: whether they depend on patients' own responses or the observations of others: and whether they are condition-specific or generic measures of health.

Some of the oldest health measurement instruments, such as the Karnofsky Index for patients with cancer (developed in the 1940s) and various Activities of Daily Living scales (developed in the 1950s), are still used today (75.161 ,503). What most of these measures have in common is that the assessment of the patient's health is usually performed by someone who observes the patient, often a clinician.

Recentl y, however, there has been an explosion of research interest in measures that incorporate the patient's own self-assessment. In particular, the past decade has seen increasing interest in the use of measures of self-assessed health status that might be applied across a wide variety of health conditions to evaluate the effects of health care technologies (6,5 19).

One reason for the surge in interest is the discovery that clinical markers of health often correlate very poorly with the patient's perception of his or her health status. Perhaps the best documented example of this phenomenon is the evaluation of the health status and progress of persons with benign prostatic hyperplasia (BPH), a noncancerous enlargement of the prostate gland. BPH is very common in elderly men and often results in a narrowing of the urethra (the urinary conduit), producing troublesome symptoms such as frequent urination and difficulty starting urination. To evaluate patients with BPH, urologists traditionally have used a measure of the amount of urine left in the bladder after voiding. They have also used a measure of the rate of urine flow to assess obstruction, and they have measured the size of the prostate through palpation and imaging. None of these measures. however, correlates well with how patients experience symptoms, or even with the frequency of their symptoms (2,25,43, 557).

A second reason for the interest in self-reported measures of health is the increasing evidence that health professionals are often not good proxies for their patients when it comes to reporting symp-

toms and health experiences (63,587). Studies comparing self-reports with reports from proxies suggest that the less observable a characteristic is (e.g., personal values about health care), the less likely it is that others can report on that characteristic accurately (227,482,655,773).

Interest in developing generic measures of self-assessed health status has derived in part from the desire to assess changes in a person's well-being when that person has multiple health conditions, and treatment for one condition can affect others (263). Generic measures also enable researchers to avoid reinventing new measures for every health condition. One widely used instrument developed for the Medical Outcomes Study, for example, has since been used in studies of such varying conditions as diabetes and knee replacement surgery (41 2,558).

A third reason for interest in generic measures of health is the desire to make comparisons across different conditions and treatments for the purposes of health policy and resource allocation decisions (59 1). This use of health status measures has been an especially strong incentive for the development of measures whose results can be summarized in a single number and incorporated into cost-effectiveness analyses (see chapter 5).

### *Measuring Health-Related Quality of Life*

Most instruments used to measure health-related quality of life take the form of questionnaires that ask about at least four different dimensions of this attribute (503,726). These are:

1. *Functional ability.* This component relates to what people can do, without regard to their resources or the actual demands on them. Physical abilities included in a questionnaire might include climbing stairs, or being able to read a newspaper or hold a pen.
2. *Perceived health.* Worry about one's health and satisfaction with one's health are commonly measured aspects of self-perceived health. Or, a question may simply ask people to rate how healthy they think they are.
3. *Psychological well-being.* This component focuses on the extent to which people see themselves as distressed (e.g., depressed or anxious). It is intended to be broader than specific measures of mental health. although it is related.
4. *Role functioning.* Questions regarding role function ask about individuals' work, their resources, and what they ordinarily expect themselves to do on a day-to-day basis (e.g., care for one self and family, visit friends), These questions help accommodate the fact that the same condition can have very different effects on people. A knee injury that severely limits the normal activities of a professional athlete, for instance. may be much less limiting to a professional editor, even though both of them have the same absolute functional abilities.

For many of the generic health status measurement instruments, results are summarized by describing the results for each of the dimensions the instrument measures. For example, a conclusion might be that a patient improved in physical function but was unchanged with regard to role functioning or perceived health. The "SF-36" and the Sickness Impact Profile (box 3-1 ) are probably the best known U.S. examples of generic instruments measuring patients' self-assessed health.

When the purpose of the measurement is to make comparisons across conditions. however, researchers instead sometimes use an instrument that produces a summary value for quality of life-one that combines results for the different dimensions measured and presents them as a single number. To come up with such a summary value, scores for the individual dimensions must be combined. usually by assigning weights to the individual scores and adding these weighted scores. The weights, which are intended to represent the relative importance of the different aspects of health being measured, might be derived from statistical models or average ratings of health care workers, patients, or the general public. The Quality of Well Being (QWB) Scale (see box 3-1 ) is among the best known examples of instruments that produce a single quantitative score of health-related quality of life.

**BOX 3-1: Examples of Instruments To Measure Health-Related Quality of Life**

Many different quality-of-life instruments exist (75,503), and the emphasis on development and use of particular measures varies among different countries. The Health Utilities Index, developed by Statistics Canada, is being used in Ontario, Canada as a general population health status measure as well as a clinical and policy tool (263). The Nottingham Health Profile has been used widely in the United Kingdom (360), and the EuroQol index has been used in a 14-country study in Europe (228) Three multidimensional measurement instruments have been particularly widely used in the United States for studies of health outcomes and effectiveness: the Sickness Impact Profile (SIP), the Medical Outcomes Study 36-item short-form health survey (SF-36), and the Quality of Well-Being (QWB) Scale

The SIP was developed in the 1970s to create a comprehensive instrument to measure the impact of sickness on people (49). Containing 136 questions that measure health in 12 different areas, it is considered one of the most comprehensive measures of health. Portions of the SIP have been used in studies of patients with conditions as diverse as pneumonia and chronic pain (328,396) and by some of the federally funded Patient Outcomes Research Teams (PORTS) (263).

The SF-36 comprises 36 questions about 8 different aspects of health-related quality of life (892) Its great strength is parsimony, it is fairly brief to administer while capturing most of the information obtained from much longer surveys Like the SIP, the SF-36 has been well-studied, and the instrument (or portions of it) has been applied by the PORTS and in other research on a wide variety of conditions (263)

The QWB differs from the SIP and the SF-36 in that it was specifically designed to produce a single score that represents an Individual's reduction from perfect health (414). This instrument consists of a list of questions that ask the respondent to report opinions or experiences regarding various symptoms (e g , headaches), diagnoses (e.g., blindness), and activity limitations (e. g , being unable to drive a car) Respondents' answers are individually weighted according to the relative importance of those problems (based on pre-existing preference weights derived from surveys conducted by these researchers) and then totaled to produce the overall score The QWB has been used both in clinical studies of outcomes (413) and for health policy purposes, in the development of Oregon's prioritized list of Medicaid benefits (788)

SOURCE Off Ice of Technology Assessment, 1994, based on sources as shown Full citations are at the end of the report

### *Applications anti Limitations*

Disease-specific health measurement tools are standbys of health research, both because of their sensitivity to the nuances of the health condition of interest and because they are often designed to be simple and inexpensive to administer (590). New disease- and condition-specific tools continue to be developed and validated, and many emphasize patients' self-assessments (42,103,486, 656).

Generic tools such as the SIP and SF-36 have the great advantage of enabling standardization and comparability of results across multiple conditions studied (590). Brief versions of such generic measures offer the possibility of much more detailed monitoring and comparisons of the health status of specific populations (e.g., enrollees in particular medical practice or health insurance plans) than is now possible (50). The greater use of generic measures in clinical trials could add

to the understanding of the relative benefits of competing medical technologies and enhance clinicians' and patients' abilities to make informed decisions about treatment choices (316,847,848, 852).

There is growing agreement that generic measures of self-reported health status and quality of life can be reliable and valid for both health status monitoring and for comparing the outcomes of specific therapies. There is also growing experience in aulgmenting them with disease-specific questions to make them more sensitive. Current research is focused on: which measures are best for which applications: how and when to use disease-specific measures, or adapt general measures for specific diseases: how answers to these questionnaires might differ across specific subpopulations: and how to minimize the number of questions that must be asked while still capturing the essence (263,420,590).

There is some debate, however, about how best to encourage clinical researchers to incorporate quality-of-life measures into trials. Staff at the National Institutes of Health point out that adding this component can increase the cost and complexity of trials, and that trial researchers may resist incorporating it (849). AHCPR officials and advisors, on their part, express frustration at the sense that this component is frequent] y considered an external add-on, with experts in the topic consulted well after a trial has been designed, rather that including it as an intrinsic part of a trial (127,816).

Some of the difficulty in getting quality-of-life measures incorporated more extensively into clinical trials may derive from the fact that the superiority of these measures over existing measures of health outcomes is not clear to the trial researchers. Head-to-head comparisons of existing trial outcome measures with generic quality-of-life measures might be required to demonstrate the superiority (or lack of it) of the generic measures.

The most controversial area regarding the use of quality-of-life measures is the application of the findings from quality-of-life surveys to health policy and resource allocation decisions. Unlike applications to clinical outcomes studies, this use of health measurement tools essentially assumes that the average scores from instruments such as the QWB represent the value that society as a whole places on different levels of health. This issue is discussed in more detail in chapter 5.

## Primary Studies To Evaluate Health Effects

Epidemiological studies to observe and compare the health outcomes of patients are the backbone of medical science. They can be roughly divided into two categories: observational studies, in which the actual experiences of the groups being compared are simply observed: and experimental studies, in which the experiences of the groups are intentionally influenced by the study.

*Observational studies* are the traditional source of information on suggestive associations in epidemiology. The recent reports of a series of similar cases of fatal and near-fatal illness among Native Americans in the rural southwest, for example. has suggested the introduction of a new infectious disease (832). Case-control and cohort study designs are types of observational studies commonly used to make direct comparisons where experimental designs are infeasible.[2]

In *experimental studies,* study participants are randomly allocated among treatment and control groups. Random allocation is intended to ensure that all comparison groups are reasonably similar not only with regard to known characteristics but also any characteristics that are unknown but

[2] In case-control studies, a group of individuals with the characteristic of interest ( cases ) are compared with individuals without that characteristic (controls) regarding their previous exposure to some factor. In cohort studies, individuals are classified according to their exposure or nonexposure to a disease or intervention and followed for\\ ard to track the outcomes.

might influence the outcome. Differences in the outcomes of the groups thus can be attributed to differences in the treatment, with a level of confidence that can be described statistically. In general, the larger the groups, the greater the level of confidence that an observed effect truly exists and was not merely due to chance.

Where the effect of an intervention is large and immediate, evidence based on the findings of nonrandomized observational studies is often enough to draw a confident conclusion that the effect is at least real. '*Slam-bang" technologies such as blood transfusions and antibiotics were convincing because the outcomes after the interventions were so dramatic compared with the expected course of the conditions they were used to treat.

But most modern medical advances are incremental rather than revolutionary. They are aimed at such improvements as reducing disease complications in people with diabetes (162), or slowing the decline in the cognitive functioning of people with Alzheimer's disease (153,230). Conversely, the predicted benefit of a new technology (e.g., immunotherapy treatment for women with recurrent spontaneous abortion) sometimes turns out to be illusory (267). In these circumstances, the ability to reliably distinguish real but modest effects from no effects through carefully performed studies is crucial to the credibility of the conclusion.

Randomized studies maximize internal validity—the certainty that (he treatment actually caused the effect. Because they are specifically designed to disprove the null hypothesis (i.e., that the treatment has no effect), they err on the side of finding no apparent effect even where a very small one actually exists. In contrast, nonrandomized study designs tend to favor the treatment being tested (784). Where both randomized and nonrandomized control studies of a new therapy have been performed, the nonrandomized studies generally (although not always) find the new therapy to be much more beneficial than do the randomized ones ( 136,529,669,950).

Thus, the validity of nonrandomized studies that conclude that a particular technology is beneficial is often suspect, because of the known bias in favor of finding a beneficial effect, and because it is often impossible to assess the extent to which the groups being compared were actually similar. This problem is especially acute in "case series" studies, where the "control" is how well patients have done in the past, or how well the case patient was expected to do in the absence of the new treatment. The hope generated from apparently positive results inferred from case series can make the inability of later randomized studies to show any effect especially disappointing (box 3-2).

Randomized study designs are unquestionably superior in being able to link cause and effect. No other tool offers the ability to exclude extraneous explanations with such confidence. Although for ethical and logistical reasons a randomized controlled trial (RCT) is not always possible, it is nonetheless well establishcd as the method of choice (88,390,784).

Despite their advantages. the RCT study design is frequently criticized as a basis for drawing conclusions about the effectiveness of medical technologies, Some of the major criticisms are:

- The applications of most technologies have never been tested in RCTs (779). Therefore, if decisions are needed now, other evidence must be used.
- Randomized trials of accepted technologies are difficult to conduct and may be unethical, because many physicians and patients already believe these technologies to be effective.[5]
- Some particular types of interventions. such as psychotherapy and new surgical techniques, have posed challenges to randomized study designs (see e.g., reference 730). Innovations in these areas e.g., laparoscopic surgery) often

[3] Peto has argued that it is equally unethical to prescribe untested therapies to patients even if these therapies are common (604).

## BOX 3-2: Lorenzo's Oil

The story of "Lorenzo's oil," a combination of two fatty acids purified from olive oil, has been widely publicized, particularly with the release of a 1993 film dramatizing the efforts of Lorenzo Odone's parents to find a cure for their son's illness It was the Odones who first hypothesized that the mixture might be a therapy for adrenoleukodystrophy (ALD), a rare disease that causes the degeneration of myelln, the protective covering of nerve fibers Although there is also an adult form of ALD, the disease usually strikes boys between the ages of 5 and 10, resulting in death wlthln a few years of onset

In 1984, Augusto and Mlchaela Odone resorted to studying ALD themselves after doctors told them that Lorenzo's illness was untreatable. After extensive research, the Odones tried treating their son with what has become known as Lorenzo's 011, they became convinced that the mixture not only halted the progress of ALD but also caused a partial reversal in Lorenzo's condition

The Odones challenged the medical community to validate their accomplishment with formal clinical trials Despite high hopes for Lorenzos oil, it has not been proven to be effective once the identifying neurologic symptoms of ALD appear in boys (640,641) Also, a recent clinical trial to test Lorenzo's oil for sufferers of the adult form of ALD failed to yield evidence that it was effective (29) Although physicians have been prescribing Lorenzo's 011 for several years in hopes that it will hold off the disease in young boys who have not yet developed the symptoms of ALD (577), its therapeutic value is now thought to be much *more* limited than first suggested (640)

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

gain acceptance before they can be identified and studied by those outside the immediate practitioner community.

- The expense and administrative difficulties of establishing and running randomized trials, and the delay before answers are available, makes it impractical to conduct RCTs on every use of every technology.
- Trials are frequently too small to detect any but the largest effects, rendering a finding of "no effect found" difficult to interpret. Although the U.S. research establishment, and particularly the Department of Veterans Affairs (VA) and NIH, has considerable experience in collaborative efforts to conduct very large clinical trials, such trials to date have generally also been expensive.
- Strict trial protocols intended to ensure that any effect found can be attributed to the treatment being studied have often limited the generalizability of RCTs. In the past, for example, most trials of therapies for acute myocardial infarction (heart attack) excluded elderly persons in order to avoid any confounding due to the comorbidities that many elderly persons have (373). Women of childbearing age have also systematically been excluded from many clini - cal trials on the grounds that some women might be pregnant and the technologies being tested might prove harmful to the fetus (524). The consequence is that the results of many trials cannot be applied with confidence to women or to elderly persons.
- The fact that clinical trials are often conducted in teaching hospitals, by specialists, on highly selected patients according to strict protocols. makes their conclusions suspect when applied by community physicians to their patients in other settings. The surgeon performing a procedure in a clinical trial, for example, may be much more skilled than other surgeons who will later carry out the procedure (335 ). Pa-

tients who participate in trials are often more motivated or less sick than patients not participating in the trials (434).

A recently completed RCT of intensive insulin therapy for persons with diabetes exemplifies some of these criticisms. The trial successfully confirmed that intensive therapy yielded benefits beyond those of standard therapy ( 162). However, the intensive therapy regimen requires constant attention and commitment by the patient and has some risks; in the trial, patients were highly motivated, received much clinical attention, and were not representative of the general diabetic population. Although this very expensive trial certainly established that intensive therapy was more efficacious than standard therapy, clinicians are expressing doubts that its findings have much practical use or are attainable under ordinary circumstances (64,619).

As a consequence of these perceived barriers to randomized trials, and the fact that many medical innovations are not subject to regulatory review of their evidence of effectiveness, most interventions never undergo evaluation with RCTs (779). Furthermore, those that do often have not been evaluated with respect to the full range of patients and practitioners that use them. Two responses to this situation have emerged. One response, which has been emphasized by the federal government effectiveness research program, relies on innovative ways to assemble and study the observational data that currently exist in administrative health databases. The other response relies on innovative ways of applying the basic principles of RCTs to make them more adaptable to community settings or to a wider range of interventions. These innovations are described below.

### *Database Studies*

Disillusioned with the lack of useful, relevant information existing for many medical technologies, many health policy makers and researchers embraced ideas for enhanced research use of large health care administrative databases in the late 1980s (247,295,373). Medicare and other health insurers keep computerized records of patient claims, which include such information as patient age, sex, diagnosis, procedures performed, and the charges billed for those services. Advocates of the greater exploitation of these administrative databases as research resources pointed out that observational data from these sources have a number of advantages compared with the collection of data through RCTs. These include:

- the large numbers of patients represented in the data;
- the fact that the data represent ordinary medical practice, rather than very selective patient populations or settings;
- the immediate availability of the data;
- the ability to track patients' health experiences back over time;
- the unobtrusiveness of data collection; and
- the expectation that analyzing existing data should be much cheaper than planning and implementing entire new trials.

On the other hand, researchers such as Byar have criticized this research technique, on the grounds that if the purpose is to compare medical technologies, administrative databases—like all observational data sources--contain biases that often render the results invalid (94,95,299,747, 893). The heavy role these databases play as sources of information in effectiveness research warrants a detailed examination of their uses.

#### Descriptive uses of administrative databases

The data from insurance claims and hospital discharge databases have long been used to describe various aspects of health care. Medicare claims data, for example, have been used widely to provide estimates of the costs of health care for elderly persons and to examine the characteristics of persons who incur high costs (26,954). Analyses of the direct medical costs of specific illnesses have also drawn heavily on information from administrative data (35,47 1,639).

The use of administrative data by researchers in the 1970s and early 1980s to describe tremendous variations in medical practice patterns across different areas and populations focused attention on the potential power of this tool. The use of admin -

istrative databases for documenting variations in medical practice continues to draw considerable interest, because of the implications of these variations. Administrative data from several sources, for example. have been used in studies that found that African Americans undergo coronary artery bypass surgery at lower rates than do white Americans (287,921,931). Such findings can stimulate a search for ways to improve access to services for particular populations. The Institute of Medicine (IOM) suggests that the existence of documented practice variation is an important criterion for the selection of medical technologies to assess, because it implies that uncertainty or disagreement exists in the field (377). The IOM has also suggested that the existence of such uncertainty improves the likelihood that an assessment can affect clinical practice, although this assertion is open to challenge (see chapters 7 and 8).

Administrative databases have also been useful for describing trends in the use of individual medical technologies of interest. Some studies have used this tool to monitor trends in a technology's applications over time, such as the increasing use of total hip replacement surgery (647). Researchers have also used administrative data to describe changes in the treatment of prostate cancer, and to examine whether guidelines intended to affect treatment patterns were associated with any changes (690).

In a related application, administrative data have proven useful in describing the relationships between new and existing technologies. One recent study, for example, found that a new technique to improve blood flow to the legs—peripheral artery angioplasty—was associated with an increase rather than a decrease in peripheral artery bypass surgery, an older technique with a similar purpose (767).

Examining the health outcomes (e.g., mortality rates) associated with the use of particular procedures has been one of the most publicized uses of administrative databases. Studies of mortality following transurethral prostatectomy and carotid endarterectomy (650,939) have been quoted widely. Published studies using administrative data have also examined the impact of specific procedures on outcomes such as rehospitalization (24) and reoperation (648).

Descriptive studies making use of administrative databases encounter a number of generic problems related to the data sources. One of the most pervasive issues is whether the numerically encoded diagnoses and procedures that appear in administrative databases accurately represent the real circumstances involving that patient. Some well-recognized problems include inaccurately assigned codes, particularly when coding accuracy does not affect payment (158.249,358,365): incomplete codes, particularly for patients with multiple diagnoses and procedures (5 14); and difficulty ascertaining whether a coded condition was actually a pre-existing condition or a consequence of treatment (929).

A second generic issue for studies using admin - istrative data is the actual difference between the population represented in the database and the population of interest. For example, a study describing rates of a procedure among veterans that used administrative data from the Department of Veterans Affairs health care system might underestimate procedure use, since these data would not capture procedures performed in non-VA hospitals.

Describing practice pattern variation, trends in the use of particular technologies, and health outcomes associated with particular technologies and patterns of care is relatively straightforward. The validity of the description depends largely on the extent to which the database examined. and the analysis of it, was appropriate to the question. The researcher must be confident, for example, t hat the database actually represents the entire population of interest, and that the occasions where the technology was applied are reasonably complete and accurately recorded.

But when the descriptive information derived from the database is used to suggest associations between trends and events, new issues arise. Like any such conclusions based on observational data, these arc always subject to a healthy dose of skep-

ticism, because the researcher can never know that the event of interest+. g., a change in technology or the release of a guideline—was the sole, or even a major, cause of the observed trend. Confidence in the conclusion can be strengthened by showing that no other known relevant events occurred, but it rests on negative evidence.

The major caveat that suggestive associations are only that—suggestive—is especially true for one current use of administrative databases: their use in documenting and contrasting health outcomes resulting from a particular procedure or medical intervention. The fundamental task of documenting health outcomes is a purely descriptive one. Once documented, however, those outcomes are frequently used, either implicitly or explicitly, to compare the outcomes—and, by inference, the relative effectiveness—of alternative medical technologies.

### Comparative uses of administrative databases

One potential comparative use of administrative databases would be to contrast the outcomes reported in a clinical trial during the investigation phase of a technology, and the outcomes that occur when that technology is in general use. Such comparisons might illuminate differences in the efficacy of a technology under strict conditions and the outcomes borne out in widespread use. It is widely believed that there are substantial differences between outcomes in randomized studies and outcomes in general practice; however, it is surprisingly difficult to find specific documented cases.

A second potential application is to compare the database-derived outcomes of apparently similar patients undergoing alternative treatments. This application, however, raises more serious issues.

As with any observational study, the valid it y of comparative results derived from information in large administrative databases rests heavily on the degree to which the populations being compared are truly equivalent in all relevant respects. Unlike randomized experiments, however, it cannot be assumed that if the groups are large enough, any material characteristics-e. g., those risk factors that make someone more or less likely to do well after a procedure—are likely to be evenly divided between groups.

As Byar observed, “in medicine, the doctor chooses the therapy precisely in order to affect the outcome” (97). Patients’ medical and other characteristics are generally expected to differ among groups receiving different therapies. To exclude these differences as reasons for different outcomes among patients receiving different therapies, researchers may examine the data to see if they can detect known risk factors. In their analysis, the researchers then “adjust” the results to account for the different distribution of these risk factors across the study populations. The degree to which the populations being compared are equivalent, and the analytic results valid, thus depends heavily on whether the researchers know all of the risk factors that might affect the results and can identify them accurately in the data.

Where identified differences in the outcomes of apparently comparable groups are very large, the differences are probably real, although the real differences may not actually be as large as the apparent ones. Probably the most striking example of this to date is a recent finding regarding outcomes in patients who have undergone cataract surgery. As a part of that procedure, some patients also undergo posterior capsulotomy, an additional optional procedure sometimes done to prevent the future development of certain visual problems. (Patients who do not undergo capsulotomy at the time of cataract surgery but who later develop the visual problems can have the procedure at that later time.) In their examination of the outcomes associated with cataract surgery, database researchers found that the rate of retinal detachment—a rare but severe complication that can lead to blindness in the eye—was over three times higher in patients who had undergone the additional procedure (39 1). Because the difference was so great, even after accounting for possible differences between patients selected for the ad-

junct procedure and other patients, the finding was a credible one.

But other outcomes comparisons have proved misleading. An early finding based on database analyses, for example, was that the mortality rates of men who had undergone traditional open surgery for prostate disease were lower than those of men who had undergone transurethral prostatectomy (650). Until then, the transurethral procedure had been considered the safer and less invasive of the two alternatives. Subsequent research confirmed that the less invasive procedure was associated with higher mortality even after adjusting for population risk factors as represented in the database (140). But this research, a more detailed review of the actual medical charts, also revealed that the patients chosen for the transurethral procedure were sicker than those chosen for the open procedure. Thus, patients undergoing the less invasive procedure probably had a higher mortality rate afterwards because physicians tended to refer sicker patients to the procedure that they perceived to be less risky. This tendency could not be identified in the original claims database, making inferences about relative effectiveness of the two procedures based solely on claims data misleading.

Unfortunately, adjusting the data to account for measures of illness severity that are represented in standard administrative databases can sometimes actually make things worse. In one study, researchers adjusted the data to account for differences in a number of secondary conditions across patients, and they discovered that analyses of the adjusted data actually suggested that patients with serious illnesses had *higher* survival rates than patients without those illnesses. At the time, the researchers speculated that the reason for this anomalous finding was the limitations to the number of diagnosis codes that could be entered on a discharge abstract (leading certain diagnoses that are often secondary to be mentioned only when there were no more important diagnoses to be coded) (395). But a later study found that the same kinds of anomalous findings occurred even when there were no restrictions on the number of codes that could be included (363).

To test more directly whether it is possible to use administrative data to make valid comparisons among technologies, several researchers have compared the results of observational studies, including administrative database analyses, with the results of clinical trials.

A recent example that is still the subject of some controversy was a study of the drug lidocaine, in which researchers hypothesized that the drug would help prevent deaths in people with myocardial infarction. An observational study found that prophylactic administration of lidocaine was beneficial in this population (356). Yet subsequent randomized trials, and a meta-analysis synthesizing those results, were unable to find any effect (346). This example is particularly interesting because the researchers in the observational study were especially careful to use stringent entry criteria, a well-defined endpoint, and adjustment for differences in risk of the endpoint. Although the trials, and even the meta-analysis, were not powerful enough to detect small differences in the subcategory of deaths most likely to be preventable with lidocaine, this comparison certainly raises the question of whether even rigorously conducted observational studies can be relied on to give valid answers.

Three important examples exist of observational studies whose results were confirmed in randomized trials. One study compared the results of a clinical trial of tonsillectomy in which children were randomized with the results of a study of children whose parents refused to participate in the trial, but who were followed observationally. The results of the randomized and nonrandomized portions of the study were indistinguishable (584).

A second study compared two technologies for coronary artery disease (coronary artery bypass grafting and medical management) using information in the Duke Database for Cardiovascular Disease. It identified people in this database

who would have been eligible for each of the previous randomized trials of this topic, and predicted what the survival for these patients would have been if they had all undergone medical management, and if they had all undergone surgery. These survival curves were then compared with the actual survival of participants in each arm of the trials. The differences between the trial results and the database analysis results were remarkably small and within the range that would have been expected simply due to random variation (349).

The third study examined the effectiveness of beta-blocking drugs after heart attack, based on observational data on patients in a particular hospital. The researchers compared observational data with results from a specific randomized trial of the drug. The results were in agreement, not only in the direction of benefit (the drugs were found effective) but in the approximate magnitude of the benefit (357).

Thus, it is clear that it is possible to obtain valid results from observational database analyses comparing technologies. It is also clear that it is possible to get invalid results. One interpretation of the lidocaine study is that "sometimes nonrandomized studies will tell you the right answer, sometimes the wrong answer, and there is no way to tell the difference without an RCT to determine the 'true' answer" (929). Unfortunately, unlike randomized trials, multiple database analyses with similar results do not necessarily raise the level of confidence that the answer is the true one, because the same unknown bias—an unknown but important risk factor, for example—may pervade all the analyses.

The finding that differences in outcomes between two groups are large lends more validity to the findings of a database analysis, because unless unknown biases are also very large it is likely that the direction, if not the magnitude, of the finding is correct. For studies in which the expected differences in outcomes are smaller, the validity of this technique for making direct comparisons is much more questionable.

There are some factors that seem to increase the chance that the results of a comparative database analysis under these circumstances will be valid. They include:

1. Detailed and pretested knowledge of the **risk factors relevant to predicting an outcome.** Hlatky and colleagues, who performed the study of observationally based versus experimentally based results in coronary artery disease management, point out that the salient risk factors for death due to coronary artery disease are well studied (929). Where modelers can predict risk of death better than most clinicians, the effect of patient selection bias in who received what therapy becomes less important.
2. Access **to sufficient clinical data in the database to detect risk factors such as secondary chronic illnesses.** The analysis of prostate surgery based solely on claims data would have been problematic even if all relevant risk factors had been known, because those factors were not adequately represented in the data.
3. **Great care in designing the database study—i.e.,** ensuring that the populations, encounters, and procedures being measured actually represent what the researchers want to measure. This factor by no means assures validity, as the lidocaine example shows, but it is difficult to believe that a study could be valid without it.

These factors exist together for relatively few conditions. In addition, these factors together still cannot guarantee validity, although they increase its likelihood. Some researchers have pointed out that even in administrative databases that are supplemented with added clinical data, it is difficult to answer questions that were not formulated carefully before data collection began (929).

Currently, considerable effort is being made to address the second of the three factors above: the need for richer databases. New directions include combining and augmenting existing databases to produce much richer sources of information. The Health Care Financing Administration (HCFA) and the National Cancer Institute, for example, are collaborating in an effort to merge Medicare claims data with cancer registry data (849). HCFA is also pilot-testing a study that will augment ex-

isting Medicare claims data with survey data on health statutes of beneficiaries (766). Combining Medicare data with data from other payers (e.g., the VA. private insurers) and with research-related data, to gain a more complete record of beneficiaries health care experience, is also an area of interest (799).

Another use of administrative and other descriptive databases, with broad potential application, is the use of a database as a sampling frame from which to draw patients for a prospective study. Medicare databases. for example, include data on nearly all elderly individuals, making a random sample drawn from it a good representation of persons in this category. Administrative data can also include information that can be used to focus the selection of individuals for a study. A study of the quality of medical care after the implementation of Medicare's prospective payment system for hospitals, for instance, used claims data to identify patients with target conditions discharged from hospitals before and after the payment system was put in place (406). Hospitals' individual computerized billing records have been used to identify appropriate persons for studies of specific conditions (702). Large administrative databases make particularly useful sampling frames for case-control studies in which cases are rare and difficult to identify through other means (396,397).

Finally, one of the most important contributions of analyses of large administrative databases may be to illuminate uncertainty and provide a focus for discussion of its resolution. This may have been the central benefit of the prostatectomy finding. Despite the fact that the actual comparative outcomes of the two prostate procedures were misleading, they highlighted the degree of uncertainty in the field. Indeed. to some degree they created it by convincing practitioners that the presumed benefits of the less invasive procedure were not obvious, and by pointing out the degree of variation in practice patterns for the two procedures.

### *Innovations in Randomized Trials: Large, Simple Trials*

One experimental technique successfully applied to overcome some of the problems of traditional randomized trials is the large, simple randomized trial. The fundamental characteristics of such a trial are:

1. it enrolls a very large number of patients. enabling it to detect even very small differences between treated groups with confidence; and
2. it is very simple in design. requiring data collection on only a few significant endpoints (604,95 1),

One of the best known trials ever conducted was essentially a large, simple trial, although it was not labeled as such. Forty years ago, the National Foundation for Infantile Paralysis recruited a team of physicians and public health researchers to mount a huge trial testing the efficacy of the Salk polio vaccine (266). In the spring and summer of 1954, the vaccine was administered to over 200,000 U.S. schoolchildren, with an additional 200,000 receiving a placebo injection. The outcome measured was simply the rate of hospitaliza - tions for polio in the test areas, Over the course of only a few months. the trial demonstrated the effectiveness of the vaccine in preventing a serious and disabling disease.

In the case of the polio vaccine, the size and simplicity of the trial design were to a great extent dictated by the urgency of the public health problem. Nonetheless, the trial remains a convincing demonstration of the potential power of the large, simple trial technique. It implemented many of the principles of this technique that have only more recently been formally described.

The modern prototype of the large, simple trial was the original ISIS (International Study of Infarct Survival) project, the first of a series of collaborative trials testing therapies to treat acute myocardial infarction (heart attack). The first ISIS trial, ISIS- 1, began in 1981 with the goal of examining the effects on mortality of the intrave-

## BOX 3-3: The Thrombolytic Trials: GISSI, ISIS, and GUSTO[1]

Following the success of ISIS-1 in determining the effectiveness of atenolol administered after myocardial infarction, cardiovascular researchers turned the spotlight on thrombolytic drugs, a major area of controversy in the field One thrombolytic drug, streptokinase, has existed for years, but until the mid-1980s it saw relatively little use The 20 randomized trials that had previously examined the efficacy of this drug had shown conflicting results, due to the fact that the effect of the drug was modest (a reduction in death from heart attack of about 10 to 30 percent) and the individual trials were fairly small, enrolling only a few hundred patients.

The first of the thrombolytic trials, GISSI-1, compared streptokinase with "usual treatment. " It recruited over 11,000 myocardial infarction patients in Italy and included as participating centers nearly 90 percent of that country's coronary care units. It ultimately documented a reduction in mortality of about 18 percent associated with the use of the drug (309).

Following the GISSI trial, ISIS-2 again used the large, simple trial design, with the participation of hospitals in 16 countries, to examine the relative effects on survival of aspirin, streptokinase, and a combination of both The trial enrolled over 17,000 patients over a three-year period and demonstrated an incremental improvement when both drugs were used in combination (385).

The emergence of new, expensive, bioengineered thrombolytic drugs on the market led both the **ISIS** and GISSI collaborative groups to conduct additional trials in the second half of the 1980s In ISIS-3, over 41,000 patients in 17 countries were randomized to a head-to-head comparison of three different thrombolytic drugs streptokinase, TPA, and APSAC.[2] In addition, half the patients

[1] The GISSI and ISIS collaborations have conducted multiple trials, which are distinguished by abbreviations and numbers (e g , GISSI-2 is the second trial conducted by the GISSI collaborators)

[2] TPA and APSAC are relatively new bioengineered drugs

nous infusion of atenolol (a beta-blocking drug) immediately after the hospitalization of a patient with suspected myocardial infarction (384). Previous evidence on beta-blocker drugs suggested a net benefit but was not conclusive. Although the expected effect of the treatment was small—perhaps a 10-or 20-percent reduction in mortality-it would be important if documented. The total number of deaths from myocardial infarction is large, so even a small reduction in mortality rates would translate into many lives saved.

To detect such a modest effect, however, the researchers calculated that they might need to enroll up to 20,000 patients (384). They formed a network of 245 hospitals in 14 countries to enroll patients. To encourage the participation of these centers in the trial, the trial organizers specifically designed the trial procedures to include very simple entry criteria, treatments, and follow-up. Entry into the trial was based on only a few specific patient characteristics, and randomization occurred over the telephone. Outcome measures were primarily in-hospital and post-hospital mortality. Ultimately, over 16,000 patients were enrolled, and the trial did indeed detect a statistical y significant reduction of about 15 percent in deaths among patients treated with the drug (384).

The success of ISIS-1 led to the use of this basic trial design in a series of additional collaborative trials examining the effectiveness of thrombolytic drugs, administered shortly after the onset of myocardial infarction to break up the blood clot (thrombus) blocking blood flow to the heart (box 3-3). The purpose of these trials was to establish,

**BOX 3-3 continued: The Thrombolytic Trials: GISSI, ISIS, and GUSTO[1]**

were assigned to receive the anticoagulant drug heparin, while the other half received a placebo (in addition to the thrombolytic drug they received) (386), Simultaneously, GISSI-2 compared streptokinase and TPA, again with a secondary test on half of all patients with heparin as adjunct therapy (310) Both trials showed that although TPA reduced later heart attacks more than streptokinase it also resulted in an Increased risk of stroke leading to an insignificant difference in overall mortality between two drugs of greatly dlffering costs

The GUSTO trial was undertaken after both ISIS-3 and GISSI-2 failed to show any slgnificant benefits from TPA over streptokinase TPA advocates hypothesized that the lack of apparent effect was due to the fact that TPA had not been administered in the most effective fashion, rapidly and in conjunction with Intravenous heparin (the previous trials had used subcutaneous heparin) (314) GUSTO did show an advantage to TPA under these circumstances (314), although the trial is still being debated in the clinical community (450,637)

ISIS-4 and GISSI-3, both completed in 1993, tested additional promising treatments for acute myocardial infarction oral nitrate, oral converting enzyme inhibitors, and Intravenous magnesium, alone and in combination They found that although enzyme inhibitors dld lower the mortality rate from heart attack, the other therapies had no clear effect (either on the overall population or on subpopulations of women and elderly persons) (31 1,381)

KEY APSAC anisoylated plasminogen-activator complex GISSI - The Gruppo Italiano per 10 Studio della Streptochinasi Nell Infarto Miocardio GUSTO Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries ISIS International Study of Infarct Survival TPA tissue-type plasminogen activator

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

first. whether such drugs do in fact reduce deaths from heart attacks; second, whether one thrombolytic drug is more effective than another: and third, whether the administrate ion of adjunct drugs such as heparin (an anticoagulant) improves the effects of thrombolytic drugs.

### Implications and limitations

The ISIS trials have provided solid support for two guiding principles of large, simple trials. First, they demonstrated that modest but clinical] y important treatment effects could indeed be detected with confidence if sample size was sufficiently largc. In the case of thrombolytic therapy for myocardial infarction, for example, even a randomized trial of 2,000 patients would not have been sufficient to reliably detect a 20-percent reduction in short-term mortality. This consideration becomes increasingly important as adjunctive therapies need to be tested—for example, heparin in addition to a thrombolytic drug, Each incremental advance in treatment can be expected to have only a modest absolute impact on an already reduced mortality rate. Head-to-head comparisons of treatments for their comparative efficacy and differences in side effects also require especially large sample sizes (92).

Second, the ISIS trials demonstrated (hat highly simplified procedures and data requirements can induce many health care providers to participate in the trial, enabling the enrollment of the large number of patients needed. The broad trial participation by providers and broad patient entry criteria, in turn, ensured that the results of these trials would have direct relevance to the broad spectrum of acute heart attack patients seen in ordinary clinical practice.

Yusuf and colleagues have argued that simplifying the trial design is not merely a poor, second-best solution where very large sample sizes are needed. They maintain that the most "important" effective treatments are those that are broadly applicable and practical (95 1). Such treatments are preferable to equivalently effective treatments that are highly complex. Because the most widely practicable treatments are often those that are fairly simple, the clinical trial protocols testing these treatments can be relatively simple, too. And simple trial protocols can be implemented without undue burden on community hospitals and practitioners, the very settings where most broadly applicable treatments take place (951).

If data collection is limited to major endpoints—those most likely to directly affect physicians' and patients' decisions regarding treatment—the trial protocol can be kept simple (95 1). In the case of thrombolytic therapy, the use of a surrogate endpoint-e. g., the destruction of the clot blocking blood flow to the heart—is actually a poor indicator of the likelihood of short-term mortality, which is the endpoint that is probably the most important to patients and their physicians.

Although patients' characteristics span a wide spectrum, and those characteristics can affect the outcome of treatment, treatment effects are nonetheless expected to be largely in the same general direction. The magnitude of the effect maybe different in patients with different characteristics, but in general all groups of patients would still be expected to have a reduction in mortality (or whatever change in major outcome is being measured) (95 1). Therefore, entry criteria in a very large trial measuring a few major endpoints can be very broad without sacrificing the validity of the results. An additional strength of trials using broad entry criteria is that their findings are of direct relevance to the broad spectrum of patients seen in ordinary clinical practice.

A very important practical strength of large, simple trials is that, because the trial protocol is kept very simple, the per-patient cost of conducting the trial can be kept relatively low. For example, in a relatively traditional clinical trial begun in 1977, the Beta-Blocker Heart Attack Trial, researchers randomized 3,800 patients at a cost of $20 million (71), or an inflation-adjusted per-patient cost of over $11,000 (92). In contrast, an ongoing trial testing the drug digitalis among patients with congestive heart failure, which employs a simplified protocol, has a total budget of $16 million and has randomized 7,790 patients, for an inflation-adjusted cost of slightly over $2,000 per patient **(92). A** large, simple trial of aspirin and beta-carotene in healthy men, the Physicians' Health Study, has had annual costs of approximately $80 per participant for the first five years of the trial (92).

A major potential limitation of large, simple trials is the other face of one of its defining characteristics: the need for simplicity in design. Corresponding to the need for simplicity is the need to collect data on only a very few critical patient characteristics and outcomes, such as mortality. The minimal data collection burden is what enables a large trial to be conducted at low cost and in community settings, but it also means that the richness of detail provided by many traditional trials is lacking. Although proponents of large, simple trials argue that most of these details are of far less importance and are therefore unnecessary to collect anyway, many U.S. researchers may be uncomfortable with their absence. A simple trial, for example, would not collect detailed information on quality of life effects, or background biochemical information from detailed and intensive laboratory tests that could be used in other aspects of research into the mechanisms of disease.

A less discussed but equally important limitation of large, simple trials in the United States is that they depend on a committed infrastructure of community health care providers, many of whom at present have had little experience in participating in clinical research. Although one of the guiding principles of large, simple trials is to minimize the number of physician encounters and tests related solely to the research protocol so that the research and data collection burden on providers is kept light, investigators at each center nonetheless

must be organized and trained. Furthermore, the providers must be reassured that patient care costs associated with the research will be covered by payers. If participating providers cannot be assured of recouping these costs, either from private health insurers or from the research sponsors, they are unlikely to stay committed to the project.

One of the benefits of creating a broad communit y research infrastructure is that it can be used repeatedly in future studies. Both the ISIS and GISSI collaborative groups have been able to perform repeated studies building off of their initial network of participating centers.

### Current applications

Despite their ability to address some of the criticisms of traditional RCTs and their impressive track record in the area of treatments for heart disease, large, simple trials are still used relatively infrequently in the United States. One of the areas where they might find potential application is in primary care and preventive services, where an intervention often must be applied to a very large population in order to see the ultimate effects on major morbidity and mortality. In fact, one of the few prominent U.S. examples of a large simple trial is the Physicians' Health Study, which is examining the preventive effects of taking low-dose aspirin and beta-carotene regularly (box 3-4).

One recently begun treatment trial that draws on some of the principles of large, simple trials is testing the use of digitalis to treat congestive heart failure, a condition in which the contractions of the heart become progressively weaker. Digitalis is a commonly prescribed and longstanding drug used to treat this condition, but the small trials conducted on this therapy have yielded inconsistent results on mortality. A large multicenter trial with a relatively simple trial protocol began in 1991 with funding from the National Heart, Lung, and Blood Institute and should provide definitive findings on the net mortality effect of digitalis (92).

Several investigators have suggested that identification of the most useful treatments for AIDS might be efficiently achieved through the conduct of large, simple trials, because most therapies are likely to have modest rather than overwhelming effects (96,98,22 1,223,707). As with the incremental addition of new therapies to treat myocardial infarction, each new therapy must show equal or greater effectiveness in relation to an expanding array of standard therapies (92). And unlike the case with many other conditions. an established network of community physicians willing to participate in research studies of AIDS treatments already exists (22 1,223).

To address the need for more extensive data on some aspects of the treatment tested, some researchers have suggested that selected participating sites (e.g., academic clinical centers) could augment the basic data collection with additional, more detailed data gathering. This strategy was used successfully in a trial of the effectiveness of routine fetal movement counting in pregnancy. which randomized over 68,000 women in a variant of the large, simple trial design (293). Although the primary outcome in the trial was a simple one—perinatal mortality—researchers also gathered more detailed data (e.g., on psychosocial effects) from a subset of the women participating in the trial. A hybrid approach such as this might be particularly appropriate for AIDS treatment, where the rapid development of new experimental treatments means there is frequently much less long-term experience with a drug's toxicity or other effects than is often the case with agents being tested in large, simple trials (92).

Researchers have also proposed that factorial designs might be productive in large trials of AIDS treatments; for example, one arm of the trial could compare two antiretroviral drugs, while another compares treatments to prevent occurrence of opportunistic infections (92). Other uses of large, simple trials in AIDS treatments might be to test the effects of different dosages of particular drugs (221 ) and to compare new antiretroviral drugs against existing therapies (707).

Even trials with very large sample sizes do not always provide unambiguous answers when the differences being measured are very small. The question of the relative effectiveness of TPA and

## BOX 3-4: The Physicians' Health Study

The Physicians' Health Study, an ongoing randomized, double-blind, placebo-controlled trial, began in 1982. The goal of the trial was to test simultaneously the effect of low-dose aspirin on cardiovascular disease and the effect of beta-carotene on cancer risk, among a population of apparently healthy U.S. male physicians. Trial participants were randomly assigned to one of four treatment groups aspirin alone, beta-carotene alone, both active agents, or both placebos.

The entire study—including dispensing study medications and collecting annual followup data on compliance and health outcomes—was conducted by mail. Physicians were chosen as study participants because they were presumed to be accurate reporters of their own health, and they were presumed more likely than other populations to comply with the necessary daily pill-taking regimen for an extended period of time. Self-reported compliance was tested in a subset of participants, and self-reported outcomes were confirmed against medical records.

The initial assembly of the study population was more involved that of many large, simple trials in disease treatment However, once participants were randomized, the trial procedures and followup in the study were highly streamlined The annual cost per participant is about $80 per year. After 10 years, compliance rates are over 80 percent, morbidity followup is over 95 percent, and no participants have been lost to mortality surveillance.

In 1988, the aspirin component of the trial was terminated prematurely due to the emergence of a statistically extreme 44 percent reduction in the risk of first myocardial infarction among those assigned to aspirin. At that time, there were insufficient numbers of total cardiovascular deaths—the trials' primary endpoint—to permit reliable assessment of aspirin's effect on this outcome, but the extremely low mortality rate precluded informative results until at least the year 2000 There were also insufficient numbers of strokes to permit reliable assessment of aspirin's effect on this outcome. Although aspirin may prevent strokes that result from blockage of arteries to the brain, because this drug increases the tendency to bleed, it is possible that its use increases the risk of the much less common, but clinically more severe, strokes that result from bleeding into the brain These questions are currently being addressed in the Women's Health Study (below).

One way in which the Physicians' Health Study diverges from the principles of large, simple trials as articulated by Yusuf et al. (951) is in its relatively homogeneous study population (male physicians aged 40 through 84). This homogeneity reduces generalizability (e. g., to women). In this case, homogeneity was purposefully selected to ensure valid results within an acceptable time frame, and the direction of effect is expected to apply to a more general population even if the exact balance of benefits and risks differs.

Low-dose aspirin in women is being tested in a separate trial, the Women's Health Study, in which approximately 40,000 female health professionals are being enrolled to evaluate the balance of risks and benefits of low-dose aspirin, beta-carotene, and vitamin E in cardiovascular disease and cancer.

SOURCE Adapted from J E Buring, M A Jonas, and C H Hennekens, "Large and Simple Randomized Trials Strengths, Successes and Limitations, " paper prepared under contract to the Off Ice of Technology Assessment, forthcoming 1994

streptokinase, two thrombolytic drugs, is still being energetically debated in the U.S. academic community despite the findings of several very large trials comparing the two drugs (see box 3-3). Controversies such as this one suggest not that RCTs are inadequate, but that some differences may be so small that factors other than relative average effectiveness must be the basis for choosing between two interventions.

### *Firms Trials*

Another intriguing application of the randomized trial design is the firms trial, in which patients are randomized among entire clinics or other institutional settings (105,541 ,564,863). Neuhauser (563) describes firms research as resting on three underlying basic concepts:

1. **Parallel providers of care.** Patients are assigned to one of several providers, who may be anything from a single physician to an entire hospital. Neuhauser points out that systematic assignment to parallel providers is not unique to firms research: existing examples include medical societies' referring patients inquiring about physicians to the next physician on their lists (to equalize referrals), and Boston City Hospital's historical assignment of new patients in rotation to the Boston University, Harvard, or Tufts teaching services (323).
2. Ongoing random assignment of patients to these parallel providers. At one hospital in Cleveland, this literally means the assignment of patients by randomly generated numbers, to ensure that the assignments are fairly distributed. Similarly, all new staff and attending (i.e., patients' personal) physicians are also randomly assigned. Once assigned, however, staff remain with their team to permit patient/provider continuity.
3. Continuous evaluation and improvement. "A change can be made in the way one provider or firm provides care, leaving the other as is. Differences in care can be observed. If the change is favorable, then it can be implemented by all firms. This becomes the new platform of care, and the next change can be started in the same way" (563). Because the firms are parallel in structure, and the patients are randomly assigned, any change in outcome can be reasonably attributed to the change in care, rather than to other patient or provider characteristics.

The idea of firms trials was first implemented at the Cleveland Metropolitan General Hospital, which began randomizing patients to care settings in 1980 (133,891). The hospital set up an experimental clinic to which relevant patients were randomly assigned when they came to the hospital for their outpatient care. Patients not referred to the experimental clinic received their care in an equivalent clinic providing usual care (564). Each clinic operated as an independent firm-hence the name of the technique.

#### Advantages

The power of firms trials is that not only changes in specific therapies but changes in the processes of care can be evaluated, taking advantage of all of the design strengths and statistical validity of randomized controlled trials. The technique is particularly amenable to studies of educational interventions and health delivery changes, and in fact many of its applications have been in those two areas.

Two major advantages of firms trials are that the basic structure is always in place, with random patient assignment happening continuously; and that the intervention itself is carried out as part of a patient's ordinary care. The consequence of these two features is that the incremental costs of conducting a study of a particular intervention are very low. Researchers cite the cost of one randomized trial testing a change in computer-based feedback to house staff, for example, as less than $1,000 (342,561). Neuhauser points out, rather colorfully, that the cost of writing up the study was the largest component cost of conducting it (563).

#### Limitations

A disadvantage of the firms approach is that hospitals or physician practices with relatively few patients cannot realistically maintain truly parallel providers and still assign enough patients to

each provider to permit statistically valid conclusions. Neuhauser notes, for example, that existing firms systems are mostly in general internal medicine, and that there are fewer pediatric care settings with enough patients to have parallel providers (563). Similarly, trials within a single institution must generally focus on common conditions or processes in order to keep sample sizes large enough for valid results (105).

Firms trials must also grapple with methodological issues that affect the validity of their results. A crucial concern is the potential for the clinics, presumed to be equivalent due to similar structure and ongoing randomization of patients and physicians, to become less equivalent over time. Problems could occur if, for example, patients in one clinic had a higher drop-out rate than another over time, or if staff had differences in expertise that was relevant to the topic of the trial (105,155).

As they are other clinical studies, cross-firm *contamination* and the *Hawthorne effect* are ongoing methodological issues in firms trials ( 105). Staff from the clinic in which the intervention is being introduced, for example, might discuss it with staff from other clinics, leading to changed behavior in the "control" clinics as well. Or, the simple fact that staff in one clinic know that an experiment is ongoing may lead them to change their behavior in ways that affect the results.

Certain kinds of trials are not well suited to firms research, at least as it is currently carried out. A clinical trial that requires the presence of a highly specialized physician, for example, would be ill-suited for this design (105).

The most significant limitation of the firms method. however, is probably in the initial difficulties of implementing a system of parallel providers with random physician assignment. Doing so requires, for example, that a randomization procedure be designed and taught to staff (559). It can also require substantial changes in the duties of individual physicians and nurses ( 194,559), which might often encounter considerable organizational resistance or require adaptations of the firms design. In one hospital, for example, the administrators discovered that it was not possible to randomize individual private physicians to firms without also assigning their partners or covering groups to the same firm. This hospital also found that the number of patients admitted overnight could vary substantially among firms as a normal consequence of random assignment, resulting in resident physicians' complaints about unequal workload ( 194).

### Applications

The number of health providers with established firms research systems is still very small, although it is no longer limited to only one or two unusual institutions. A significant newcomer, for instance, is the Wade Park Veterans Affairs Medical Center in Cleveland (439). Firms systems are in place in at least eight other hospitals, including one other VA hospital and an army medical center (564). The technique has not yet spread outside of academic medical centers, however. probably in part due to the need for ongoing expertise in such fields as biostatistics and epidemiology in order to carry out research (761 ).

Applications of the firms trial research design to date have been on quite diverse topics, with educational and service delivery topics prevalent. Examples included research on colorectal screening performed by nurse clinicians ( 104); counseling patients to quit smoking ( 134); alcohol dependency counseling (286): and the influence of new physician staffing patterns (704).

Future applications for firms research might include research into the effects (on, e.g., costs, patient health outcome, and patient and physician satisfaction) of implementing new clinical practice guidelines. Indeed, some previous trials have been conducted on subjects that examine specifically the effects of interventions to change physician behavior, Hershey and colleagues used the firms system to study the effects of computerized reminders to clinicians on practice change (341). Researchers at the Reganstrief Health Center in Indiana have conducted a series of studies that investigated the effects of various interventions on physicians' test-ordering behavior. They studied,

independently, the effects of displaying to the physician a patient's prior test results; the probability that the test would be abnormal; and the patient charges for each test ordered, and discovered that each intervention resulted in a reduction in the number of tests ordered (761).

The general concept successfully tested in firms trials—that prospective randomization based on units larger than the individual can produce valid results—is still relatively rare but has considerable potential. Medical practices, health care plans, communities, or other units are all possibilities. Bakketeig, for example, has suggested testing the effects of different ways of providing prenatal care by using geographical areas as the units of randomization (36). Such trials might be more difficult to carry out in heterogeneous countries such as the United States, but the suggestion serves as a reminder that additional innovations in the use of firms and other variations of controlled clinical research methods might be rewarding.

## Secondary Techniques To Synthesize Results

Despite its drawbacks, the medical literature is nothing if not voluminous. As Glass so succinctly stated (about the social science literature) in 1976:

> We face an abundance of information. Our problem is to find the knowledge in the information. We need methods for the order] y summarization of studies so that knowledge can be extracted from the myriad individual researches (28 l).

This section describes two different methods for synthesizing information, each with a different purpose. The first, meta-analysis, is aimed at synthesizing research results in order to draw more powerful and confident conclusions about the state of the world they describe. In its purest form it is a straightforward research tool, but it is also being used as a way of drawing together information for decisionmakers. The second tool, decision analysis, is expressly oriented to the purpose of organizing existing information and assumptions for decisionmaking.

### *Meta-Analysis and Other Systematic Reviews*

The traditional method of synthesizing the results of previous research on a topic is the research review, a discussion and analysis of work to date on the topic of interest. The need for reviews in order to make sense of existing research is great enough to support entire periodicals that publish nothing else (e.g., the publications of Annual Reviews, Inc.).

Despite science's reliance on reviews to synthesize pre-existing results, the traditional narrative review often suffers from a number of weaknesses. Reviewers often do not define clearly the methods they used to identify and select information, they often review the information haphazardly, and they rarely assess the quality of data systematically (550). The consequence is that two researchers reviewing the same topic, and even the same group of studies, can come to diametrically opposed conclusions (457). The burgeoning liter--ature and conflicting reviews have led to increasing use of more systematic reviews of the literature, using structured methods to reduce the opportunities for bias (583). A type of systematic review that has received particular attention recently is meta-analysis—a structured review that incorporates statistical methods to combine the results of the individual primary studies (220,437, 860).

The idea of combining study results quantitatively dates from 1904, when Pearson summarized the relation between inoculation against enteric fever and mortality by calculating the average correlation between those variables across five communities (593). Meta-analysis as a formal discipline, however, arose out of work on the social sciences literature in the 1970s (281,458).

The essential characteristics of a meta-analytic review are that it is systematic and quantitative (473). Meta-analysis requires that the analyst undertake a formal, explicit consideration of what literature will be represented in the review. In addition, the analyst does a quantitative reanalysis of the relevant results of those studies (box 3-5).

## BOX 3-5: The Steps of a Meta-Analysis

A meta-analysis is a systematic process (190,213,436,668) that involves nine steps:

1, Defining the research question, The analyst specifies the treatment under investigation, its alternative, the outcome, the study populations, and the quantitative effect measure of interest

2. Defining the admissibility criteria for studies, Examples of possible criteria are that for a study to be considered relevant, it must: be blinded, compare the treatment with a placebo, include elderly persons as study subjects, be written in English, and present results in such a way that the effect measure of interest can be calculated,

3. Searching for relevant studies. This step usually involves a computerized literature search, supplemented by perusing the reference lists of identified articles, abstracts from conferences, and any other informally identified sources.

4 Reviewing the retrieved studies for admissibility, The analyst reviews the identified articles to see if they meet admission criteria, abstracts relevant information, and if necessary re-expresses study results in a standard fashion for subsequent statistical analysis,

5 Assessing the quality of the admissible articles. Objective methods for assessing study quality are frequently used, and published criteria exist (90, 112, 159). Subjective criteria and criteria specially tailored to the research question under investigation have also been employed (51 ,472). Study quality criteria might include, for example, whether the investigators in the study knew which patients received treatment and which received placebo; whether the presentation of data was appropriate, and whether the statistical analyses were appropriate,

6. Correcting for probable bias. If a treatment effect observed in a given study is not an accurate measure of the true treatment effect, the study is biased Certain study designs are associated with known biases, for example, studies in which the investigators know which patients got which

(The broader term, "systematic reviews," includes meta-analyses, but it also includes reviews that undergo the same process without the quantitative step.)

In most meta-analyses, the quantitative reanalysis involves recalculating individual study results so that the treatment effects are all portrayed in a consistent manner. If some results are portrayed in the original articles as differences (between outcomes of treatment and control groups) and other results portrayed as ratios, for example, the analyst might recalculate them so they are all portrayed as ratios. In addition, the analyst must calculate, for each treatment effect, the precision with which that effect was measured in that study (i.e., the variance around the treatment effect). In general, smaller studies will have larger variances—less precision—than larger studies, because there is a greater chance that random variation is responsible for the observed effect in a small study. The analyst weights each study result according to its precision and then combines all the results in a single calculation to assess the overall treatment effect implied from the studies as a group. Because the meta-analysis of results includes many studies, it has more precision than any individual studies. Thus, a meta-analysis can increase the confidence that a real effect does (or does not) exist, even when individual studies differ in whether they find an effect.

Rather than combining the results of individual studies, some analysts actually combine the raw

**BOX 3-5 continued: The Steps of a Meta-Analysis**

treatment tend to find a larger treatment effect than do studies in which the Investigators were blinded to the subjects treatment (1 36). Where the size of the likely bias is predictable, the observed treatment effect can be adjusted to account for this bias (209,303,882) Bias correction IS often a skipped step, however, because there is often no basis for estimating the Iikely size of the bias, or even whether it exists

7 Analyzing the data The data analysis step is the one in which the results of the individual component studies are actually reanalyzed. Often the Individual results are displayed graphically A study with a large confidence Iinterval, displayed as a long bar, represents less certainty about the result Combining the results of all the individual studies, weighted by the degree of certainty of each result, gives a new result with very high confidence (i.e , a small confidence Interval) Individual studies may find no effect, but collectively the meta-analysis has the power to detect even very small effects with much greater certainty As part of the data analysis, the meta-analyst also conducts sensitivity analyses (e g , to show whether the results of the meta-analysis depend heavily on one or two particular studies)

8 Assessing publication bias The process of peer review and journal publication can winnow out studies that are considered to be less interesting simply because they found no effect, biasing the published literature in favor of studies that found effects The analyst may take steps to account for this

9 Interpreting the results As with other studies, the final step of a meta-analysis is for the analyst to interpret the results so that their generalizability and Implications for practitioners and researchers are clear

SOURCE Adapted from M P Longnecker Tools for Health Technology Assessment Meta-Analysis paper prepared under contract fo Off Ice of Technology Assessment forthcoming 1994

data from the studies. In such a "pooled analysis." the data are treated as if they are all from a single very large trial, rather than from many independent trials. Pooled analyses and meta-analyses usually give similar results. Pooled analyses facilitate the analysis of subgroups of patients, but they often require the cooperation of many scientists in order to obtain the raw data (473).

Well-done recta-analyses seem to be reasonably well established as reliable and valid. In one assessment of the reliability of this technique, Chalmers and colleagues investigated 20 replicated meta-analyses (111). They found that the differences in meta-analyses of the same research question were "almost always of degree rather than direction" ( 111 ). A similar study of meta-analysis reliability by Henry and Wilson (336) found similar results. It also found the recent meta-analyses it assessed to be generally more reliable than the older analyses studied by Chalmers et al.

In addition to comparing the results of meta-analyses with each other. both groups of researchers also compared the results found through meta-analyses with the results of single, large randomized clinical trials. Chalmers and colleagues found agreement between meta-analyses and later large trials for just one of three meta-analyses studied, while the one comparison conducted by Henry and Wilson found that the recta-analysis and the clinical trial results agreed (113,336). Other instances of agreement between individual

meta-analyses and clinical trials have also been found (442), although no comprehensive comparative survey of the field has been attempted.

There have, of course, been instances of disagreement among meta-analyses as well. Several researchers who have conducted meta-analyses of the literature on interventions to reduce cholesterol, for example, have concluded that lowering cholesterol does not result in lower mortality rates overall (150,15 1,544,621 ). In contrast, Law and colleagues concluded from their own meta-analysis that reducing cholesterol levels reduces the risk of ischemic heart disease and does not raise the risk of death from any other cause except stroke, a risk outweighed by the reduction in heart disease deaths (443). Law and his collaborators point out that the different conclusions derive in part from the different outcomes examined (all-cause mortality vs. disease-specific mortality) and differences in the trials selected for analysis.

### Issues

The growing acceptance of meta-analysis notwithstanding, meta-analytic results can be controversial (487,889). Three issues are especially hotly debated in the field: the combinability of results from the studies used in a meta-analysis, the importance of procedures to account for publication bias, and the protocol followed by the analyst.

**Combinability.** The justification of meta-analysis is based on the assumption that the component studies all address similar research questions. If the populations, the treatment, the study design, and the outcomes measured in each study are sufficiently similar, then the meta-analysis is analogous to a multicenter clinical trial. Differences in the treatment effect across the component studies—the "centers" of the meta-analysis "trial"-can be presumed largely due to chance (473).

As the component studies of the meta-analysis become less similar, the appropriateness of their joint analysis becomes a matter of judgment and is thus subject to debate. Many of the criticisms of meta-analysis revolve around this specific aspect of the technique (69,251.274,352). For example, one meta-analysis of nonmedical treatments for chronic pain calculated the average effect of one treatment on several different kinds of pain. Some studies included in the meta-analysis measured headache pain, while others measured cancer-related pain. Critics of this meta-analysis charged that the treatment effect might have been very different for headache and cancer pain (352). If this were the case, summarizing across the different types of pain might have obscured the true treatment effects in these different groups.

An equally controversial issue in combinability is whether it is appropriate to combine the results of nonrandomized studies. While evidence from good randomized clinical trials is widely accepted as valid, the validity of results from nonrandomized trials is less clear, and these results are excluded from many meta-analyses (602). Some researchers define meta-analysis to include only analyses of randomized studies (91,928).

For many research questions, however, only data from observational studies are available (46). Dickersin and Berlin (170) point out that a meta-analysis of such studies should be as acceptable as are the studies themselves. The crucial point is that the meta-analysis cannot entirely overcome the deficiencies of the studies on which it draws: if the studies are biased, the results of the meta-anal - ysis will probably be biased, too.

Separating analyses of studies based on study design may be one way of detecting and reducing potential bias. In a meta-analysis of alcohol consumption in relation to risk of breast cancer, the estimate of the effect of alcohol derived from the combined case-control studies was larger than the estimate derived from the combined follow-up studies (472). The authors kept the analyses of the two types of studies separate and argued that, for various reasons, the result based on the combined follow-up studies was the more like] y to represent an unbiased result.

A third issue in combinability of studies arises when the treatment effect found in the component studies varies markedly among studies. Summarizing a single treatment effect across studies under these circumstances is commonly done, but

when and how to do it are subjects of debate among researchers (303,603).

A common quantitative method for combining study results is the "random effects model." in which the calculated summary treatment effect is assumed to be an estimate of the average treatment effect in the universe of hypothetical studies with differing treatment effects. The component studies in the meta-analysis are assumed to be a sample from this universe. However, some analysts prefer the "fixed effects model," which assumes that there is a single "real" treatment effect that the different component studies are all attempting to estimate, with varying degrees of success.

In practice. the two methods give similar results when the results of the component studies of a meta-analysis are not too variable. When the results of the component studies do vary substantially, the "'fixed effects" model gives heavy weight to the largest studies, while the "random effects" model gives a result somewhat closer to a simple average (473).

**Publication bias.** *Publication bias* refers to the well-documented fact that studies that get published differ from studies that do not. in ways that are not just related to the quality of the study. Several researchers have shown. for example, that studies with statistically significant results are more likely to be published than other studies (46, 170, 193a). Results perceived as important are also disproportionately likely to be published (172,193a).

One of the characteristics that sets meta-analy - ses and other systematic reviews apart from traditional narrative literature reviews is the use of explicit rules for including studies in the analysis, and researchers in the field of meta-analysis have carried on a longstanding debate about how to prevent, or adjust for, publication bias. A number of formal statistical methods to detect and assess the extent of publication bias in a meta-analysis have been proposed, but as yet there is no widespread agreement on their use (473).

Some researchers suggest that the solution to this problem is to include all relevant unpublished studies, as well as the published ones, in the meta-analysis (269,952). Most analysts agree that when unpublished studies can be obtained, they should be assessed along with published ones ( 143). Differences in the results of published and unpublished studies can be assessed by presenting the results of the meta-analysis with and without the unpublished studies (143). Unless registries of all studies undertaken in a given field exist, however. including all unpublished studies may be impractical or impossible (171,890,952).

**Meta-analytic protocol.** In addition to procedures for summarizing treatment effects and for accounting for publication bias, meta-analysis researchers debate a number of other aspects of the meta-analytic process.

Chalmers, for example, argues that the evaluation of studies to be included in the meta-analysis should be blinded (107). He follows a protocol in which the names of the authors, the actual results of the studies, and other study characteristics that might bias the reviewer are hidden during the study selection process. In addition, he recommends that two people independently evaluate the quality of the studies in a meta-analysis (107). Most researchers agree that these procedures should improve the quality of the meta-analysis (473). They come at considerable cost in reviewer time, however, and the degree to which they improve the quality of the analysis has not been shown. Hence, they are often not followed.

Considerable debate also surrounds the issue of how best to judge the quality of the individual studies considered for inclusion in the meta-analysis. One possible option, for example, is to assign each study a numerical score according to how well it meets each of a number of specified indicators of presumed quality ( 159). Low-scoring studies could be excluded, given a lower weight in the analysis, or analyzed as subgroups. Alternatively, Rubin has suggested that characteristics of component studies be analyzed in relation to the treatment effect, to see if particular characteristics (e,g., study design) strongly affect the result of thc meta-analysis (662). There is no uniform protocol

or agreement among analysts regarding the approach to follow. There is agreement, however, that explicit attention to study quality is important (473).

Finally, some researchers specifically advocate a Bayesian approach to meta-analysis (210,288). This approach explicitly incorporates the analyst's own presumptions about the likelihood of certain things, such as whether a particular study to be included might be biased. Its potential advantages include statistical results that are easier to interpret than those of traditional meta-analysis, and greater flexibility in combining different types of information in the meta-analysis. Its disadvantages include the need for special software to perform the analyses, the greater susceptibility of the results to debate (because the analyst's assumptions are fundamental components of the analysis), and the fact that even fewer people understand Bayesian methods than understand traditional meta-analysis (473).

### Applications

Meta-analysis is unquestionably gaining in popularity, application, and influence, especially in medicine and public health. The number of published meta-analyses on health topics, and articles about meta-analysis, has grown from fewer than 100 in the entire decade prior to 1987 (171) to over 200 in 1989 and well over 300 in 1991 alone (473). Topics range from the usefulness of prophylactic antibiotics for children with recurrent ear infections (933) to the effect of garlic on cholesterol levels (895).

Evidence from meta-analyses has been used to support a number of the federal government's clinical practice guidelines. Because it not only synthesizes existing information but adds value to it, by a more robust estimate of whether a given health care intervention is effective, meta-analysis has become a standard input to the Agency for Health Care Policy and Research's guidelines effort (80 1,802,8 10). The U.S. Preventive Services Task Force also considers meta-analyses as evidence for its recommendations (87 1). According to a member of the Task Force, a meta-analysis is given the same grade of evidence as the grade that would have been applied to its component studies (868). The U.S. Food and Drug Administration allows the results of meta-analyses to help support new drug applications (25 1).

The U.S. General Accounting Office has proposed that meta-analyses be conducted that combine results from randomized clinical trials with those from analyses of large administrative and other databases (882). The purpose of such "cross-design syntheses" is to enable statements about a treatment's effect in the general population (that represented in the database) to be made, while grounding the certainty that the treatment is efficacious in the randomized trial data. Since the essence of this method is a meta-analysis that combines randomized with observational data, it is likely to be controversial, and its validity may be difficult to establish. The feasibility y of the technique is currently being tested by GAO researchers (700).

Recent research suggests that while individual trial populations may differ from the population at large, pooling the results from many trials may give a more representative finding. Klawansky and colleagues examined age-specific survival rates in four clinical trials of breast cancer patients and compared them with U.S. cancer registry data (430). They found wide variability in survival rates across trials, suggesting that individual trials did indeed vary from each other and the general population of breast cancer patients. When the results of the trials were pooled, however, the overall survival rates were quite similar to average survival rates for those age groups in the registry. Thus, the problem of nonrepresentative trial populations may be lessened if the results of multiple trials are combined.

In summary, meta-analysis' applications in areas where multiple randomized trials exist are considerable. Under these circumstances, the technique permits a statement about two treatments' relative effects to be made with considerably more certainty than is possible from the individual trials, and it is a useful tool in assessing effectiveness. It can enable more robust estimates

not only of the efficacy of an intervention in a broader population than is enrolled in any one trial, but also on particular subgroups of special interest (e. g., elderly persons) to see if there are differences in effectiveness for those subgroups. Its major limitations are, first, that relatively few researchers are trained in the technique; and second. that the reliability and validity of a meta-analysis—indeed. the ability to do one at all—are limited by the studies that exist for it to draw upon.

### *Decision Analysis*

Decision analysis, a technique for guiding rational decisionmaking under uncertainty (620), has been rapidly gaining in its application in health care effectiveness research and technology assessment. 1( is not a new field, nor is it historically associated with health care, but its applications in this area are spreading rapidly, and some of the implications of those applications have considerable public policy consequences.

The essence of a decision analysis is the systematic. schematic presentation and examination of all of the relevant information for a decision, the points at which decisions or uncertain events occur, and the relative preferences the decisionmaker would have for the array of various possible outcomes of the decision. A simple decision analysis is frequently depicted as a decision '-tree," which branches at points of decision (e.g., surgery vs. no surgery) or uncertainty (e.g., getting a post-surgical infection vs. no infection). (See box 7-4, p. 162, for an example.) The decision analyst records, at each appropriate branch, the best estimate of the probabilities that various outcomes might occur and what those outcomes are.

The use of decision analysis to improve medical decisionmaking was proposed by Lusted in 1971 (478). One of its most familiar (although not necessarily most frequent) health care applications has been decisions about the best course of treatment for a particular individual patient, In this context. decision analysis is primarily a way of laying out the options available to a physician or patient and organizing the information relevant to those options in a way that helps the individual make the decision. It serves as much as a discussion tool as a decision tool.

A physician, for example. can discuss with a patient that person's "preferences" for various possible outcomes of treatment. The patient in this example might assign death a "preference" weight of O, permanent disability a weight of 80, and eventual full health a weight of 100. A decision tree can then be drawn that included the various treatment options and the chances, under each option, that each of those three outcomes would occur. The physician and patient then can multiply the probabilities by the outcomes and arrive at a number representing the net "desirability" of choosing each option.

Applied to decisions for or by groups, some of the characteristics of decision analysis have additional implications. Matchar (496) argues that the greatest benefit of decision analysis as a tool to aid in expert group decisionmaking is its ability to be "a language for the representation of difficult decisions." Unlike many more complex models that can be used to aid decisions (e.g., detailed computer-based simulation models), in a simple decision analysis the information and assumptions can be laid out in a way that makes them easily comprehensible to all members of a group. The group can then discuss the assumptions and the factual information and save their most heated discussions for discussing the importance of relative outcomes, rather than on what assumptions are implicit in the model.

In addition to providing a framework and language for discussion among individuals in a group, decision analysis helps make clear what important information is missing (496). By testing the sensitivity of the results of the analysis to different estimates of preferences for specific outcomes, the group can examine the range of potential implications of its decisions.

The critical controversy over decision analysis, however, relates not to its use in organizing information but to its explicit incorporation of preferences. Calculating which decision path is preferred requires that the decision analyst assign to each possible path not only the outcome of that

**BOX 3-6: The Creation and Evolution of the Patient Outcomes Research Teams**

The concept of multidisciplinary research teams to study the outcomes of ordinary patient care pre-dates the establishment of the Agency for Health Care Policy and Research (AHCPR). The research program supporting such teams (originally labeled "Patient Outcome Assessment Research Program" grants) was funded by the National Center for Health Services Research (NCHSR) under its Outcomes Research Program The program formally began soliciting grant applications for research teams in 1988 (839a)

The assessment teams to be funded under the program were modeled on the original prostate disease outcome research team (see chapter 2 text) Each team was to focus on a particular medical condition, They were to be composed of 5 to 7 full-time-equivalent professionals and were required to include persons with expertise in at least nine specified subject areas:

- clinical competence in the study subject,
- epidemiology,
- biostatistics,
- research design,
- economics,
- decision analysis,
- survey research,
- data management, and
- research synthesis and meta-analysis (839a).

The assessment teams were also given a very specific charge as to how they should go about their research efforts (839a). They were to:

- conduct literature reviews and research syntheses of the condition,
- use existing "routine" databases to develop hypotheses about practice variation,

path but also the relative preference for each outcome. Incorporation of these preferences, or "utilities, " is part of the reason that decision analysis can be a powerful tool. But unlike the example with the individual physician and patient making a decision discussed above, assigning utilities to outcomes in the context of a decision with group-level application requires two crucial assumptions. First, it requires that the preferences being included in the analysis are the correct ones for the decision-e. g., the preferences of the group that the decision will affect. This requirement is simple to state but can be—and has been—the source of considerable disagreement in practice. Second, for the assigned utilities to be valid in the context of the decision they must be truly valid measures of the real preferences for that outcome. This topic is an area of intense empirical research and theoretical debate (see chapter 5).

Thus, decision analysis, although it depends on existing information, has several uses in clinical evaluation. It can be used in both research and clinical practice to display outcome probabilities and preferences of individual patients. It can be used in cost-effectiveness analysis, where cost as well as health outcomes are incorporated into the analysis. And it can be used in clinical practice

**BOX 3-6 continued: The Creation and Evolution of the Patient Outcomes Research Teams**

- develop more extensive data sets to examine these hypotheses, includlig the use of primary data gathering through interwews and surveys,
- based on this information, design "carefully-focused epidemiologic or experimental clinical trials that NCHSR WiII consider conducting, "
- disseminate research findings to physicians, and
- evaluate the impact of the research and dissemination on physician behavior and practice patterns

The first four teams, funded in the fall of 1989, addressed cataracts, myocardial infarction, prostate disease, and back pain (797)

AHCPR, which replaced NCHSR in 1989, funded more outcome research teams in the following years These "Patient Outcomes Research Teams" (PORTS) were to become the centerpiece of the Federal government's effectiveness initiative As with the pre-AHCPR teams, PORTS were required to conduct literature reviews and synthesis, analyze practice variations and associated patient outcomes, using available data augmented by primary data collection where desired, disseminate research fundings, and evaluate the effects of dissemination (797) By October 1992, a total of 14 PORTS (including the first four) were receiving AHCPR funding (817) No new awards were made in 1993

A recent program announcement re-inviting applications for new PORTS relaxed substantially these methodological requirements placed on the first set of PORTS Teams are still to be interdisciplinary and focus on a specific condition or problem, but they are given more leeway to define for themselves the methods they choose to address the issue (81 1 ) Six new "PORT-II ' grants were awarded under the revised program in 1994

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

guidelines development to structure expert group dicisionmaking. The lattcr two uses are discussed further in chapters 5 and 7.

## APPLYING THE TOOLS: THE PATIENT OUTCOMES RESEARCH TEAMS

The PORTS are the "showcase investment" of the federal government's effort to apply the tools of effectiveness research (800). The government's goal for them was explicit and ambitious:

> The goals of a PORT project are to identify and analyze the outcomes and costs of current alternative practice patterns in order to determine the best treatment strategy and to develop and test methods for reducing inappropriate variations (797).

The characteristics and methods of PORTS were closely prescribed, by statute and by the terms of the requests for grant applications put forth by AHCPR. In structure. PORTS were to be multidisciplinary, multi site. large-scale, and long-term (800) (box 3-6). All were required, regardless of their particular topic and thrust, to include four components: a comprehensive literature review and synthesis (e.g.. a recta-analysis); an analysis of variations in medical practice and associated patient outcomes (using claims and other sources of data); dissemination of findings about effectivc care: and an evaluation of the effects of dissemination ("to demonstrate methods that encourage voluntary change in provider behavior"). The effectiveness of dissemination was

to be judged "in terms of reduced variation in practice patterns, more appropriate use of health care resources, and improvements in patient outcomes" (797).

Cross-cutting methodological issues faced by the PORTS are discussed in periodic meetings of interPORT work groups. These groups grew out of a meeting held shortly after the first four PORT grants were awarded in 1989 (485). They offer chances for PORT investigators to explore common issues and problems and to consult with additional experts about those issues. AHCPR provides formal support for the six groups. which are on the topics of:

- literature review and meta-analysis,
- use of claims data,
- decision modeling,
- outcomes assessment (e.g., measuring quality of life),
- cost of care, and
- dissemination of findings (485).

In addition to their roles in research and information dissemination:

> the clinical recommendations developed by PORTS [were] intended to be a primary source of scientific information for use by independent expert panels in the eventual development of practice guidelines (800).

The agency has several times deliberately assigned the same medical condition to both a PORT and a guidelines panel. This dual attention is in part the result of the priority AHCPR staff have placed for both activities on high-frequency procedures and conditions that have correspondingly high costs to the Medicare program. It also has allowed the guidelines panel to take advantage of previous or concurrent work done by PORT teams. The cataract panel, for example, relied extensively on the review performed by investigators on the cataract PORT team (724a). In that case, the principal investigator of the PORT was also the consulting methodologist to the guideline panel.

In another case, one of the investigators of the prostate disease PORT was actually appointed a member of the guideline panel on the same topic. The influence of the PORT's work is evident in the emphasis the practice guideline ultimately put on eliciting patient preferences as a crucial determinant of the most effective and appropriate treatment (819).

Although a number of PORTS are much too new to be expected to have any findings yet, it is not too soon for preliminary judgments about what can be expected from this centerpiece of federal effectiveness research. Of the 14 PORTS ongoing as of early 1994, four were in the fifth and final year of their grants (table 3-1 ). Another seven were in their fourth year, the year they were to begin disseminating the results of their research. The contributions of the PORTS thus far can be judged on three grounds:

1. For the PORTS nearing completion, have the original goals of these projects been met?
2. Aside from those goals, have the PORTS contributed new insights, knowledge, or evidence regarding the effectiveness and cost-effectiveness of medical interventions?
3. Has the work of the PORTS contributed to the infrastructure of health research in other ways (e.g., through advances in methodological techniques)?

## Contributions

The PORTS have developed topic-specific expertise in great detail, using the talents of investigators with diverse backgrounds and training. Many of the methodological developments described earlier in this chapter have been in part the contributions of PORTS, particularly in the areas of meta-analysis, administrative database analysis, and the application of measures of patient functioning and quality of life. These contributions are illustrated by some of the specific output of the initial four PORTS (which have had the most time

**TABLE 3-1: Current and Planned Patient Outcomes Research Team Projects (PORTS) as of July 1994**

| Start date | End date | Topic |
|---|---|---|
| **9/89** | **8194** | Back Pain Outcome Assessment Team<br>*Uhversity of Washington, Seattle, WA* |
| **9/89** | **8/94** | Consequences of Variation in Treatment for Acute Myocardial InfarctIon (AM I)<br>*Harvard Medical School, Boston, MA* |
| **9/89** | **9/94** | Variations in Cataract Management Patient and Economic Outcomes<br>*Johns Hopkins University Bait/more, MD* |
| **9/89** | **8/94** | Assessing Therapies for Benign Prostatic Hypertrophy and Localized Prostate Cancer<br>*Dartmouth College, Hanover, NH* |
| 4/90 | **3/95** | Assessing and Improving Outcomes Total Knee Replacements<br>*Indiana University Indianapolis, IN* |
| 6/90 | **9/95** | Variations in the Management and Outcomes of Diabetes<br>New *England Medical Center, Boston, MA* |
| 7/90 | **8/95** | Outcome Assessment Program in Ischemic Heart Disease<br>*Duke University Durham, NC* |
| 8/90 | **8/95** | Outcome Assessment in Patients with Biliary Tract Disease<br>*University of Pennsylvania, Philadelphia, PA* |
| 9/90 | **9/95** | Analysis of Practices Hip Fracture Repair and Osteoarthritis<br>*University of Mary/and, Baltlmore, MD* |
| 9/90 | **9/95** | Assessment of the Variations and Outcomes of Pneumonia<br>*University of Pittsburgh, Pittsburgh, PA* |
| 9/90 | **9/95** | Variations in Management of Childbirth and Patient Outcomes<br>*The Rand Corporation, Santa Monica, CA* |
| 8/91 | **8/96** | Secondary and Tertiary Prevention of Stroke<br>*Duke Universlty Medical Center, Durham, NC* |
| 9/92 | **9/97** | Schizophrenia Patient Outcomes Research Team<br>*University of Maryland, Baltimore, MD* |
| 9/92 | **9/97** | Patient Outcomes Research Team (PORT) on Low Birthweight in Minority and High-Risk Women<br>*University of Alabama, Blrmingham, AL.* |
| 9/94 | **8/99** | PORT-II For Prostate Disease<br>*Massachusetts General Hospital, Boston, MA* |
| 9/94 | **9/99** | Cure, Costs and Outcomes of Local Breast Cancer<br>*Georgetown University Washington, DC* |
| 8/94 | **7/99** | Cardiac Arrhythmia PORT<br>*Stanford University Palo Alto, CA* |
| 7/94 | **6/97** | Homemade Cereal-Based Dehydration Therapy<br>*Health and Hospitals of the City of Boston, Boston, MA* |
| 7/94 | **6/99** | Dialysis Care: Choices, Outcomes, Costs and Tradeoffs<br>*Johns Hopkins University Baltimore, MD* |
| 9/94 | **8/98** | **Value of Medical Testing** Prior to Cataract Surgery<br>*Johns Hopkins University Baltimore, MD* |

SOURCE U S Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research Rockville MD 1994

to obtain results) and, to a lesser extent, by more recent PORTS.

## Prostate Disease

The PORT on prostate disease is in some ways the most defensible one on which to base conclusions about the contributions of this organizational form of research, because the research team itself actually predated the formation of AHCPR. It was the prototype for the PORT concept, and it has actually had several additional years to carry out its line of research. **The clearest and most widely acknowledged contribution of this team has been its investigation into the role of patient preferences and functional outcomes in treatment decisions for prostate disease (both BPH and prostate cancer).** Several of the insights into the importance of patients' reports discussed above, for example, are based on research by the prostate PORT. Among its specific contributions are:

- highlighting disagreements among physicians in treatment for prostate disease, and identifying the discrepancies between the great increases over time in the number of prostatectomies performed, and the lack of evidence that this treatment was more effective than alternatives (253,480,896,91 1);
- demonstrating the importance of patient self-assessments and preferences in determining the appropriate treatments for BPH and prostate cancer, and the discrepancies between patients' reports, physicians' assessments, and outcomes of treatments reported in the literature for these diseases (42,253,264); and
- convincing both the clinical research and the practicing urology communities that good clinical studies comparing alternative intervention strategies for BPH and prostate cancer are needed.

### *Back Pain*

**The prime success of the back pain PORT has been to demonstrate, repeatedly and convincingly, that a major reason for great variation in treatments for back pain is the utter lack of evidence that any one treatment is more effective than any other.** In one study, for example, the researchers identified a sevenfold variation in the rate of cervical spine surgery to treat neck pain among counties in the state of Washington. The authors pointed out that this large variability in practice is not at all surprising in light of the lack of clinical evidence that might support any unified approach to the treatment of this problem. The abysmal state of the literature on both the diagnosis and treatment of back pain, and the great need for good studies, is a major theme in a number of publications by the investigators in this PORT (350,769,770).

In other contributions, an interesting physician survey conducted by the back pain PORT demonstrated great variation in the diagnostic tests used for low back pain and showed that the physicians' specialty (e.g., neurology, rheumatology) is strongly associated with the type of diagnostic test ordered (11 9). PORT researchers have also worked with Maine physicians to conduct a prospective study examining the outcomes of disk herniation and stenosis (821).

### *Acute Myocardial infarction (AMI]*

In stark contrast to the back pain PORT, the AMI PORT focused its investigations in an area in which claims data were relatively plentiful and in which data from high-quality comparative studies already existed. **The major contributions of this PORT were its various examinations of the concordance between the evidence regarding effective interventions that exists, and the extent to which those interventions are applied in practice.**

Some of the most powerful findings of this PORT came from its meta-analytic studies and the contrast between treatments shown to be effective based on meta-analyses and their acceptance in the medical community (27,442). The concept and insights possible from the technique of cumulative meta-analysis were a clear contribution of this PORT to the methodological development of the field. Analyses of claims and other administrative databases also proved illuminating; they documented great variations in the rate with which

generally effective interventions are performed across gender, age, and racial subgroups (33.34,772). Although these analyses could not fully identify the reasons for these differences, they raised clear questions about whether the processes by which treatment decisions are made are fully equitable.

In conjunction with database analyses that link differences in particular treatments with mortality outcomes. AM I PORT researchers have applied some novel statistical techniques (hierarchical modeling and instrumental variable analyses) (5 15). These techniques have not yet been applied, evaluated, and critiqued by peers in detail, however, so their full contribution towards drawing conclusions about the comparative effectiveness of different technologies cannot yet be assessed.

### *Cataracts*

The PORT examining the effects of cataract surgery in Medicare patients, like the AMI PORT, had the advantage of being able to identify relevant patients and procedures in claims data with fair accuracy. Researchers examined mortality outcomes of elder-l y cataract patients overall (734) and more specific clinical outcomes associated with particular types of procedures (101 ,392, 393). For the most part. this research confirmed previous studies and estimates of complications and outcomes. Other relevant contributions of this PORT have been estimates of the costs of the episode of care surrounding cataract surgery, and a measure of vision function for cataract patients (821).

The chief success of the cataract PORT was its finding, based on claims data analysis, that a particular adjunct procedure (laser capsulotomy) maybe associated with a greatly elevated risk of retinal detachment (391). This complication is a severe one, and although the absolute risk found in the study is small, if confirmed it would imply that performing capsulotomy as a preventive procedure is not necessarily a good idea. The finding is notable because the rarity of the complication would make it very difficult to detect in a clinical setting, and because the magnitude of the increased risk makes it very difficult to dismiss the finding out of hand as an artifact of the method.

Although this is probably the most direct and credible finding of comparative safety and effectiveness based on claims data analysis (from the PORTS or other research), it has not gone unchallenged. Among the criticisms, for example. is the fact that the data do not permit researchers to identify whether the eye suffering retintal detachment was actually the eye that underwent the procedure in question. For this and other reasons, many ophthalmologists apparently do not find the results convincing (724). AHCPR is currently funding a case-control study, conducted by the same researchers, to confirm the results of the claims data analysis (724,821 ).

### *Other PORTS*

Although the longest of the remaining PORTS have been in existence for only four years, several have reported findings.

- The stroke PORT has reported differences among racial groups in the receipt of technologies to diagnose and treat the disease (575). It has also done extensive work examining the factors that predict outcomes after stroke (511 ) and the usefulness of decision models in helping expert panels rate the appropriateness of indications for carotid endarterectomy (a major surgical procedure sometimes performed to prevent strokes) (576).
- The PORT studying knee replacement surgery has reviewed the rating systems used in assessing outcomes, with the goal of helping develop and encourage more consistent and valid methods of assessing patients' levels of improvement after surgery (179). They have also confirmed that most patients do consider themselves better off after surgery, are conducting a cohort study of surgery for arthritis of the knee, and are examining the comparative outcomes of surgery in different subpopulations of elderly patients undergoing the procedure (764).
- Members of the pneumonia PORT conducted a prospective follow-up study in which they doc-

umented substantial variations in lengths of hospital stay for pneumonia patients, particularly in low-risk cases (238). PORT researchers have also developed a pneumonia-specific prognostic index that they believe could be a useful tool for clinicians (239).

## Limitations and Frustrations

Despite their several notable successes, the PORTS have suffered equally notable disappointments. Most of these are directly related to the limitations of the methods they employed for the objectives they were ostensibly trying to address.

First. **the PORTS have been largely unsuccessful at identifying the most effective treatments among the alternative treatment patterns in existence,** one of the fundamental stated goals for this research investment. The closest successes have probably been the retinal detachment finding of the cataract PORT, the findings regarding the importance of patient preferences in determining treatment appropriateness in the prostate PORT, and the possibility that the techniques of the AMI PORT might produce some credible findings on relative effectiveness of treatments. Nor do there appear to be critical findings on relative effectiveness on the immediate horizon from any of the newer PORTS. It appears to be rare that data from existing claims and other databases, even augmented ones, are sufficiently clear and show differences of a sufficiently large magnitude to be useful in drawing conclusions about relative effectiveness.

Second, **no particular research method has proven universally fruitful; the mandate to use particular research methods has frequently led to inefficient or unproductive lines of research for individual PORTS.** Exhaustive reviews of the literature, for example, have proven a very expensive undertaking for some PORTS. with relatively little to show except to document the poor quality of existing evidence (485,807). Similarly, analyzing variations and outcomes from claims data did not prove particularly useful or productive in some PORTS. In the hip fracture PORT, for example, claims data were not very useful, since nearly all patients with fractured hips are hospitalized, and since when hip surgery (or resurgery) is performed it is not possible to tell which hip was the subject of the operation (807).

Third, **the ability of the PORTS to undertake active dissemination of their findings, and to evaluate their effects on clinical practice patterns, has been very limited.** The first and most obvious reason for this is that the PORTS have offered very little in the way of delineation between appropriate and inappropriate and ineffective practices. Most of the contributions of the PORTS have dealt with insights that can improve the processes of care-e. g., through accommodating more explicitly patients' preferences and leading providers to question the equity of patients' referral to specific treatments-rather than insights into which practices lead to better health outcomes overall.

A second reason is that disseminating information, and setting up a process for evaluating the effects of that dissemination, is a very different activity than the initial research, requiring both new planning and new skills. Overall, the PORT investigators have been dissatisfied (and rightly so) with their ability to perform this function within the constraints of their five-year grants (807). Only the back pain and prostate disease PORTS appear to have fully operative dissemination and evaluation programs, and neither has yet formally presented any results from these studies.

Fourth, **perhaps the most glaring failure of the federal government's investment in the PORTS has been the inability to follow up the detailed examinations of the poor and conflicting evidence justifying current alternative practices with primary research to resolve the questions.** Both the prostate disease and the back pain PORTS, for example, identified specific questions about treatment effectiveness that could only be answered in prospective studies (9 18). For the most part, however, these studies have not been forthcoming. Exceptions include cohort studies by the back pain and knee PORTS, the case-control study now underway to confirm the findings of the cataract PORT, and two prostate

disease studies that examine some aspects of the questions raised by the prostate PORT. In contrast to these small exceptions. the uncertainties demonstrated by these PORTS in current practice was enormous. evidence to resolve them was lacking, and explicit attempts by the prostate PORT to initiate a trial directly testing specific questions it raised were unsuccessful. This issue is discussed in greater detail in chapter 4.

## Future Plans

Although there has been no formal agency assessment of AHCPR's effectiveness research program in general, or the PORTS in particular, the agency has engaged in some introspective discussion about this approach. AHCPR held a small conference in 1993, at which agency staff, PORT investigators, and other attendees discussed some of the lessons of the PORTS and promising directions for future research teams (807). Among the conclusions of participants were:

- The approaches used by the PORTS had not been universally successful, particularly the overemphasis on claims data analysis and exhaustive literature reviews. More future emphasis on more flexibility in methods, and more small prospective studies (within the limitations of AHCPR's resources) was warranted,
- Evaluating the dissemination of their findings was not something most teams could accomplish within their time. expertise, and financial constraints.
- There was considerable merit to having interdisciplinary research teams gain indepth knowledge and expertise in a particular clinical condition, and for the more successful of the PORTS there might be merit in maintaining these centers of expertise.

AHCPR did not fund any new PORTS in the fall of 1993. Instead. the agency released a new request for application for future PORTS that was great] y changed from the initial PORT solicitation five years earlier and incorporated many of the sentiments expressed at the conference. Among the most notable differences, the new grant announcement stressed the following:

- Conditions affecting mainly children. youth, or nonelderly adults would be given as much priority as conditions affecting the predominantly elderly Medicare population.
- Investigators were encouraged to “design new research strategies, to use new combinations of methods. or to tailor existing methods” in order to obtain evidence for the comparative effectiveness of clinical interventions. Experimental and quasi-experimental research designs were explicitly mentioned.
- The use of secondary sources of data, such as claims data, was not given prominence, and suggestions for the use of administrative data were the more modest possibilities of using them “in identifying cases and controls, estimating costs, or measuring selected outcome s.”
- Researchers were specifically instructed to include women and minorities in study populations.
- There was no mandate that the research teams disseminate their findings and study the effects of that dissemination on changes in clinical practice (811).

The first six new PORT grants (PORT-II) began in the summer of 1994 (table 3-1). Some of these new PORTS follow up, with prospective studies, questions raised by the initial PORTS (300,81 7).

## CONCLUSIONS

The ability of patients, providers, and payers to get valid and reliable information on which health care technologies work best, for whom, and under what circumstances, has always been limited. AHCPR was created in 1989 in part to fulfill this need. The establishment of that agency marked not only a commitment to effectiveness research but also an emphasis on particular facets of that research. Thanks in part to the stimulus provided by the federal government’s emphasis on increasing that understanding, the tools now available to enhance our understanding are many and are continuing to be developed and refined. Their applications are likewise growing. In its level of sophistication. the science of evaluating the com-

parative effectiveness of existing health care interventions has passed from infancy to somewhere in early childhood.

**The focus on evaluating the outcomes of health interventions that most matter to patients, and the refinement of tools to achieve this aim, is one of the major contributions of effectiveness research as it has been carried out thus far through the federal effectiveness research initiative.** The current debates over which health survey questions and instruments to use for this purpose have not been resolved. Global measures enhance comparability of results across studies, while disease-specific measures offer more opportunity for relevant detail. Brief measures are simpler to use and might enable information on patient functioning and qua] it y of life to be incorporated in studies more widely, while longer measures offer more ability to be sensitive to specific problems. Areas still in need of attention are:

- Continued methodological research into different measures, and different applications of those measures, to understand more fully the advantages and drawbacks of each.
- Development of common measures for the sake of enabling more valid comparisons across studies of the same disease.
- Better collaboration between quality-of-life researchers and researchers conducting comparative clinical studies, so that study results can be more meaningful to more patients. AHCPR and NIH both clearly have much to contribute, yet cross-fertilization between the agencies on this topic has been limited. Many institutes seem to have relatively little interest in the methodological work done at AHCPR; and where there is interest, AHCPR seems to perceive it as interest in that agency's resources rather than real interest in intellectual collaboration.

The analysis of large administrative databases—a tool deliberately emphasized in the federal effectiveness initiative and the mission of the PORTs—has proved quite useful for several specific purposes. Among its important contributions have been its uses in:

- highlighting variations in medical practices and paving the way for serious discussion about the reasons for these differences, including future prospective studies;
- identifying appropriate candidates for primary studies;
- highlighting the differences between medical practices shown to be effective and their use in particular populations of patients, as demonstrated by the research involving data on AMI patients;
- identifying rare adverse events; and
- enriching clinical with administrative data, which offers possibilities for much richer descriptive information on the experiences of patients with particular conditions and undergoing particular treatments.

In contrast, **the notion that the analysis of large administrative databases could address the need for information on the comparative effectiveness of alternative treatments has proved misguided. No** clear, wholly credible finding about the direct effectiveness of one medical practice over another has been derived directly from this research method thus far. Even the finding of the cataract PORT regarding the risk of retinal detachment has not been entirely convincing to clinicians. Other research suggests that there are areas where credible, or at least highly suggestive, findings might be forthcoming, but administrative databases themselves are not the most productive means for determining the comparative effectiveness of most medical technologies and services. Focusing on this research method as a relatively simple, inexpensive first-line tool for answering comparative questions is unwarranted.

**Prospective comparative studies, and particularly randomized controlled trials, have been underused in the government's effectiveness initiative.** Although they are often considered to be tools applied to medical technologies at an early stage, variations of the randomized clinical trial design can and have been applied to compare two or more existing interventions, and to include broadly representative populations. One of the main contributions of administrative

database analysis (o effectiveness research, in fact, has been to highlight uncertainties and—even more importantly-create an environment in which patients and clinicians alike can agree that comparative trials are needed.

An aspect of comparative effectiveness trials largely uncommented on, in either the literature or the health policy debate, is the relationship between effectiveness trials conducted within a committed infrastructure and the goals of continuous quality improvement—a topic very much the subject of current discussion. This relationship is particularly marked in the GISSI large, simple trials. which included most of the coronary care units in Italy. As the trials were completed, units could incorporate the findings, and new trials were begun to achieve the next level of quality improvement. Questions of generalizability of findings were almost irrelevant. since most units and patients participated. Firms trials have accomplished this objective on an institute-specific basis; as an intervention proved effective, it was adopted by the other firms in the institutions and became the new level against which future improvements would be measured.

Large, simple trials seem a particular y promising tool for comparing the effectiveness of some interventions in ordinary practice, where large sample sizes might be necessary but provider participation and funding may require that protocols be kept very simple. The major drawback of this simplicity is that it conflicts with the need to better measure patient-centered outcomes and preferences. Further innovations in trial design might overcome this and other drawbacks+. g., through "nesting" a smaller trial with more detailed data collection in a larger, simpler trial. At least one successful example of such a "-nested" trial in primary care already exists (293). **In general, the experiences of effectiveness research thus far suggest that it is not the rejection of randomized controlled trials but innovations in the design of clinical trials, and greater incorporation of RCTs into ordinary practice, that is needed to improve the level of knowledge about the comparative effectiveness of existing medical interventions.**

**The refinement and greater application of meta-analysis and other systematic reviews of the literature is a useful contribution of the effectiveness initiative.** The experience of the PORTS suggests that while an insistence on exhaustive literature collection can be both inefficient and unnecessary, systematic reviews nonetheless have been important in highlighting both research and practice deficiencies. When good studies do exist, meta-analysis can also often derive more powerful and convincing statements from the findings of previous research. The greater use of systematic reviews could reduce unnecessary and duplicative research, enable important information already in the literature to gain broader exposure, and reduce inconsistencies among literature reviews. Although the application of meta-analysis to nonrandomized studies has limitations, it also has promise. It would behoove clinicians and health policy makers alike to learn to be able to judge the quality of a meta-analysis at a basic level, and to be able to interpret its results.

Decision analysis is a tool for organizing existing information, incorporating information on effectiveness and outcomes from pre-existing studies and structuring it to help clarify the choices to be made. Like meta-analysis, it can also help point to needed areas of research, where information to make a decision is especially poor and especially important. The power of decision analysis derives from its ability to structure the information needed to make a decision and assess consequences. Distinguishing among the quality of different kinds of studies and other information used in the analysis, however, is the responsibility of the analyst—a feature the users of decision analyses must bear in mind when using decision analysis to compare the outcomes of different alternatives.

The PORTS have been a successful testing ground for developing and applying many of the tools of effectiveness research. They have espe-

cially excelled at raising the level of discussion about what is known, and what is not, about the effectiveness of treating particular diseases. They have also contributed to an improved set of measures for assessing the outcomes of therapies for problems such as prostate disease and knee conditions.

In so doing, the PORTS have created a fertile environment for new research on existing medical technologies and services. Attempts to generate new evidence regarding effectiveness using the tools they have emphasized in the past, however, have met with only rare success and point to the limitations of a research model that, at least until now, has emphasized secondary research methods and the use of existing rather than newly generated data.

The inability of the federal government effectiveness research efforts to follow up the questions highlighted by PORT research with comparative clinical trials is one of the signal failures of those efforts. Although the next round of PORTS may include a better balance of research methods, including more comparative prospective studies, these will still be constrained by numbers and resources in the questions they can address. The implications of this and some of the other issues raised by effectiveness research thus far are discussed in the next chapter.

# Issues in Improving Effectiveness Research

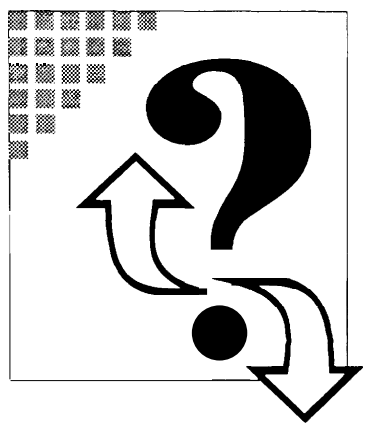
# 4

The federal effectiveness research effort has gone far in raising questions of the comparative effectiveness of existing technologies and strategies to manage health problems. It has been less successful at answering the questions it has raised.

This chapter addresses the issue of how the federal government can improve effectiveness research. To do so, it first discusses some of the major gaps in effectiveness research as it is currently carried out, and the barriers and possibilities in filling these needs. It then reviews the part the various federal agencies and departments play in this research effort, and how the roles of the different agencies affect the implementation of strategies to address problems in the current effort.

## GAPS IN THE EXISTING FEDERAL RESEARCH EFFORT

As described in chapters 2 and 3, **the effectiveness research activities sponsored by the federal government have yielded valuable insights about the relationships between the outcomes and processes of care, but they have been less successful at making clear statements about the relative effectiveness of alternative medical technologies and services.** Among the clear gaps in the existing federal effectiveness research are:



1. **The lack of a systematic assessment of what has already been studied.** Despite the enormous and ever-increasing size of the medical literature, exhaustive reviews of past studies in areas such as treating back pain have sometimes found almost nothing useful. In some cases, however, systematic reviews have demonstrated the existence of unrecognized but relevant studies. A coordinated means of assessing the results of past

studies could help ensure that useless duplication of extensive reviews are reduced, while making better use of knowledge from past studies.

2. **The absence of valid comparative studies of existing technologies.** Effectiveness research has proven adept at raising appropriate qucs-(ions to study and fostering a climate conducive to comparative clinical research on existing medical technologies, but research to address these questions has until now received little real support or commitment from federal agencies.
3. **The inability to prevent the problem of poor evidence from accumulating.** The fast pace of biomedical research, and the relatively small proportion of new technologies ever exposed to rigorous testing before introduction, mean that our collective ignorance about the most effective technologies and strategies may be growing rather than declining.

Each of these areas raises its own issues and possibilities, discussed below.

## Systematic Reviews: Making Use of Existing Knowledge

Making the most efficient use of health research resources requires first knowing what has already been studied. Sometimes our lack of knowledge regarding the safety and effectiveness of technologies is due not to a lack of studies but to our lack of awareness about them. The true tragedy of diethylstilbestrol (DES), described in chapter 2 (box 2-1 ), is not only that it ultimately proved very harmful but that it was never effective, and that its ineffectiveness could have been known from the beginning if clinicians had heeded the results of the more rigorous studies of the drug. Even some of DES harmful effects could have been detected, had contemporary analysts examined the results of their own studies more critically (106).

Systematic reviews of the literature, including meta-analysis, have proved to be a powerful tool of effectiveness research. ] The contributions of systematic reviews are threefold. First, they have encouraged a more rigorous approach to defining and conducting a literature search and review than was the norm in the past, making reviews more reliable and providing a needed tool for managing the enormous size of the medical literature. Second, they have added strength to existing evidence, and sometimes added new findings to the existing evidence, through the quantitative reanalysis of previous research results. Third, they can demonstrate areas in which the existing literature is especially weak, an important criterion in targeting resources toward the research questions most in need of investigation.

A powerful demonstration of both the need for systematic reviews and the contributions they can make was a set of meta-analyses by Lau, Antman, and their colleagues, who examined the results of published trials of treatments for acute myocardial infarction (27,442). Their findings implied that thousands of lives have been lost because physicians did not know of, or did not believe the results of, studies that had already been done. Streptokinase, for example, was little used until the late 1980s, when the introduction of a higher priced, genetically engineered alternative kindled new interest in this older drug. Early trials of streptokinase were small and had contradictory results. These researchers showed that had the results of these small studies been combined in a meta-analysis, clinicians could have known by the end of the 1970s that streptokinase, administered soon after a heart attack, saved lives. Yet as late as 1984, most major textbooks and reviews of the field made no mention of the therapy, or argued against its use. Conversely, lidocaine is still being advocated as routine therapy in textbooks, even though 20 years ago a meta-analysis of published trials

[1] "Systematic review" here encompasses both meta-analysis and other comprehensive, highly structured literature reviews that are not able to combine quantitatively the results of individual studies (e.g., because the outcomes measured are too different).

would have raised major questions about its effectiveness (27).

Despite the rising popularity of meta-analyses, the conduct of systematic reviews is also often a frustrating, inefficient, and disjointed exercise. Agency for Health Care Policy and Research (AHCPR) has funded meta-analyses primarily through its Patient Outcomes Research Teams (PORTS), with a few additional methodological studies also receiving grant funding. The PORT reviewers were frequently frustrated with the considerable resources spent on extensive literature collection that nonetheless resulted in few useful studies (807). This experience suggests that better ways of identifying the relevant literature would be a great efficiency. Furthermore, it suggests that recording and updating such reviews where they have been done could prevent others from duplicating the effort.

It might be presumed that areas in which considerable randomonized controlled trial (RCT) activity is being undertaken would be promising areas for systematic reviews of previous trials. NIH conducts and sponsors many clinical trials, but it sponsors few formal research overviews or meta-analyses and almost no methodological activities on this topic. The National Institute for Child Health and Human Development (NICHD) does have one project to study the use of meta-analytic techniques for combining the results of nonrandomized studies, and it and a few other Institutes have one or two meta-analyses that are ongoing or recently completed, but this activity receives little emphasis overall.

Nor is it likely that many researchers conduct a formal meta-analysis with their own resources before proposing a clinical study. Those who have conducted meta-analyses report that the commitment required in terms of researcher expertise and researcher and computer time can be substantial (473). Encouraging the production of systematic reviews thus is likely to require at least some external support, as well as collaboration among a number of investigators.

**In summary, meta-analyses and other systematic reviews are ways of making better use of existing knowledge, gaining new knowledge, identifying important questions for future research, and preventing the squandering of resources on previously researched questions. Such systematic reviews can be costly but at present have few sources of federal funding, and there is little to encourage researchers to conduct them before undertaking new research projects. Some of the costs and potential duplication in systematic reviews could be substantially reduced if review efforts were more coordinated, and if there were better mechanisms to help reviewers identify relevant studies more systematically.**

### *The Cochrane Collaboration*

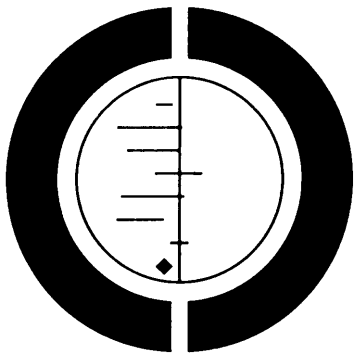
One response to the need for better understanding of what existing studies can already tell us has been the establishment of the Cochrane Collaboration, a remarkable international effort whose goal is to "prepare, maintain and disseminate systematic. up-to-date reviews of RCTs of health care, and, when RCTs are not available, reviews of the most reliable evidence from other sources" (13 1,666) (box 4-1 ).

The model for these collaborative reviews is a comprehensive review of interventions in pregnancy and childbirth, which includes systematic reviews of about 600 separate topics in the field (131). Reviewers participating in the group addressing the subject—about 30 individuals from 8 countries—prepare systematic reviews and update them as more trials on those topics are conducted.

A unique feature of the Collaboration is that the results of reviews, disseminated electronically, are not copyrighted ( 108). ". . .[A]lthough those contributing to the Collaboration are named in its electronically published output. the Cochrane Collaboration itself belongs to all of the contributors, collectively" (131).

## BOX 4-1: The Cochrane Collaboration Logo

The Cochrane Collaboration, a cooperative international network of researchers, is dedicated to the preparation, maintenance, and dissemination of systematic reviews of the effects of health care.



The Cochrane Collaboration logo illustrates a systematic review of data from seven randomized controlled trials (RCTs) Each horizontal line represents the results of one clinical trial (the shorter the line, the more certain the result), and the diamond represents their combined results, The vertical line indicates the position around which the horizontal lines had similar effects, if a horizontal line touches the vertical line, it means that particular trial found no clear difference between the treatments. The position of the diamond to the left of the vertical line indicates that the treatment studied is beneficial.

This diagram shows the results of a systematic review of RCTs of a short, inexpensive course of a corticosteroid given to women expected to give birth prematurely. The first of these RCTs was reported in 1972. The diagram summarizes the evidence that would have been revealed had the available RCTs been reviewed systematically a decade later: it Indicates strongly that corticosteroids reduce the risk of babies dying from the complications of immaturity By 1991, seven more trials had been reported, and the picture in the logo had become still stronger This treatment reduces the odds of babies of these women dying from the complications of immaturity by 30 to 50 percent.

Because no systematic review of these trials had been published until 1989, most obstetricians had not realized that the treatment was so effective. As a result, tens of thousands of premature babies have probably suffered and died unnecessarily (and cost the health services more than was necessary). This is just one of many examples of the human costs resulting from failure to perform systematic, up-to-date reviews of RCTs of health care

SOURCE The Cochrane Centre, "The Cochrane Collaboration pamphlet, Oxford, United Kingdom, 1993

Six centers around the world, including one in the United States. have been established to support the reviewers who participate in the Cochrane Collaboration (108).[2] In addition to coordinating, compiling. and disseminating the reviews in general topic areas, these centers maintain registries of systematic review's undertaken by others (131). There are no central sources of funding for either the centers or the reviewers; all of the contributors to the Collaboration are responsible for finding their own sources of support. The only U.S. Cochrane Center thus far is located in Baltimore, Maryland, It is presently subsisting on a small one-year grant from NIH's Office of Medical Applications of Research ( 169).

### *Improving the Efficiency of Systematic Reviews*

One of the most time-consuming tasks of performing a meta-analysis, or any systematic literature review. is the identification of all relevant studies to be reviewed (11 O). This task is also an inefficient one; different reviewers may each spend time trying to separately identify essentially the same studies.

The task of identifying published studies is made somewhat easier by the existence of MEDLINE®, an electronic database of the medical literature maintained by the National Library of Medicine (NLM) at NIH. Unfortunately, however. this database has several limitations that make it unreliable as a source to identify all, or even the great majority. of publishcd RCTs. Among its most prominent constraints are:

- It includes only citations to articles in the medical literature published after 1965.
- The 3,700 journals it covers represent less than 20 percent of all medical journals (and it includes few publications from related fields. such as health services research).
- The search heading used to identify RCTs (the main types of publications used in meta-analyses) was very restrictive before 1990 and did not identify the full range of trials of interest to reviewers.
- Even since 1990, RCTs are often not labeled as such on MEDLINE, because the persons entering the information on the database cannot tell easily from the published articles that they in fact are this type of study (680).

Authors can inadvertently compound the difficulties of conducting literature searches via MEDLINE. Articles in which the authors have made poor choices of key words, or have abstracts that do not clearly identify the article as an RCT, can be difficult for reviewers to identify in a MEDLINE search (865 ).

The extent of MEDLINE's limitations is demonstrated by a search of RCTs relating to vision treatments published in 66 journals in 1988. In this case, it was already known that all 66 journals were among those indexed on MEDLINE, so the retrieval rate using that database should have been very high. Of over 1,500 trials identified and examined, 201 were clearly randomized controlled trials. Another 18 turned out to be RCTs, but this was not obvious from the published articles and had to be confirmed in other ways. Of this total of 219 trials, 30 could not be identified using MEDLINE ( 168).

Literature searches of clinical trials can be even less successful if they are not restricted to recent trials published in journals known to be indexed on MEDLINE. On average, even searches conducted by an experienced medical librarian yield only about one-half of all relevant RCTs (173).

A major stride towards improving the efficiency of systematic reviews occurred in late December 1993. In a commendable example of a voluntary response to a clearly defined problem,

[2] The first Cochrane Centre was established in the United Kingdom, in O\ ford. England. The U.S. center is in Baltimore, Maryland. Other centers are 1 ocated in Canada, Denmark, 1 taly, and Australia.

NLM has committed its resources towards creating an augmented clinical trials database (173). This database will be parallel to MEDLINE, and searching MEDLINE for clinical trials will alert users to its existence. It will include:

- tags to all clinical trials already indexed on MEDLINE;
- abstracts of clinical trials published in journals held by NLM in its collection, but not currently indexed on MEDLINE; and
- clinical trials published before 1966.

The Baltimore, Maryland Cochrane Center is helping to coordinate the effort, which will result in an expanded database available to users beginning in 1995 (173).

## Filling in the Knowledge Gaps

### The State *of Comparative Effectiveness Trials*

Most of the effectiveness research sponsored as part of the federal government effectiveness research initiative has been descriptive, using administrative databases and other observational data to describe patient outcomes. A great disappointment of this research is that although it has identified important questions for comparative research, neither the funding nor the research communities have proved able to capitalize on this. Having created an environment potentially amenable to good comparative research studies, effectiveness research has been largely unable to carry out those studies. The deficiencies include comparative research on existing practices where uncertainty has been shown to exist; the broader incorporation of outcomes that measure patients' quality of life into RCTs; and more research in settings and on patients that are "ordinary," on questions that matter to patients and could further help them and their care providers make better decisions.

AHCPR has very few comparative effectiveness studies underway. The agency is contributing to a few RCTs sponsored primarily under the aegis of the Department of Veterans Affairs (VA) or NIH, and it has plans to fund a few more on its own through the PORT and other grant programs (82 1). The agency has also funded a followup case control comparative study to see whether the findings of the cataract PORT regarding retinal detachment can be confirmed (724,821). This study, and a few other small randomized and nonrandomized studies, comprise its investment in comparative effectiveness research. The agency does not consider that its current funding level permits much more than this (821 ).

NIH, in contrast, sponsors hundreds of clinical trials, but most of these are believed to be basic safety and efficacy trials of predominately new technologies. NIH does sponsor at least some comparative studies (both RCTs and nonrandomized studies) of existing technologies. Examples include:

- a comparative trial of alternative treatments for acute ear infections in children,
- a comparative trial of behavioral treatments for urinary incontinence in elderly persons,
- a study assessing the outcomes of temporal mandibular joint (TMJ) surgery,
- a large, simple effectiveness trial on the effects of digitalis on survival in patients with congestive heart failure, and
- a multicenter trial of treatment for early glaucoma (846,853).

The VA is another sponsor for a number of comparative effectiveness studies. The VA Medical Research Service's Cooperative Studies program, for example, has five large, multicenter randomized controlled trials that are ongoing or recently completed. All could be considered "effectiveness trials" in some sense, and all but one of them are cosponsored by other federal agencies. They are:

- the Prostate Cancer Intervention Versus Observation Trial (PIVOT) trial of early intervention for prostate cancer (cosponsored by AHCPR) (935),
- a trial to evaluate a new drug to reduce drug cravings in persons who are opiate dependent (co-sponsored by the National Institute on Drug Abuse) (978),
- a large trial of digitalis for heart disease (cosponsored by the National Heart, Lung and

Blood Institute and Burroughs Wellcome) (877),
- a continuing study of the role of zidovudine (AZT) in preventing progression of AIDS (co-sponsored by the U.S. Army Medical R&D Command) (876), and
- an evaluation the comparative effects of a num - ber of antihypertensive agents (497).

Neither NIH's nor VA's trials, however, are linked in any way to the priority areas for research on existing technologies that emerge from the descriptive "effectiveness research" work of AHCPR.

The inability of the existing research structure to carry out the full range of studies implied by "effectiveness research" is eloquently captured in the saga of clinical trials on treatments for benign prostatic hyperplasia (BPH), the noncancerous enlargement of the prostate (box 4-2). In this instance, descriptive effectiveness research sponsored by AHCPR raised specific questions about the relative effectiveness of common treatments for BPH. The clinical community came to accept the need for a comparative trial of the treatments and actually proposed the trial. Yet the trial went unfunded by AHCPR due to lack of money, and unfunded by NIH due to lack of interest. **One of the prime justifications for descriptive effectiveness research is to identify important research questions and illuminate medical uncertainty that would enable an RCT to take place, yet in the case where this has most clearly happened, the needed study has never materialized.**

**Improving the conduct of comparative effectiveness research requires improving the way trials on existing technologies are conducted, so that the results of the trials are as broadly applicable and as relevant to patient and clinician decisionmaking as possible.** Many improvements are possible; three that are particularly closely tied with the goals of effectiveness research are discussed in this chapter. They are incorporating broader measures of health outcomes in clinical trials, m'here relevant and possible: improving the public's knowledge of clinical trials, to broaden participation: and improving the research infrastructure so that large-scale, practice-based research becomes not only feasible but efficient.

**Equally important to improving the conduct of effectiveness trials is improving the federal government's sponsorship of such research.** Establishing high-priority questions to study, and improving the research infrastructure to study them, is useless if no federal agencies consider it one of their major responsibilities to support this infrastructure and fund research within it. This issue is discussed later in this chapter.

### *Incorporating Broader Outcome Measures*

The topic of incorporating quality-of-life assessments in clinical trials has been the subject of three separate NIH workshops (847,848,852). Despite the variety of trials in which patient functioning or quality-of-life measures are used, however, these trials probably represent a minority of NIH-sponsored trials. and the proportion apparently varies among Institutes. The National Institute for Allergies and Infectious Disease, for example, estimates that "at least 10 percent" of its trials incorporate such measures. Several other Institutes report using such measures but list only a few examples, suggesting that these measures are not major endpoints in most trials (846).

A few of these trials incorporate generic quality-of-life instruments, such as the SF-36 and the Sickness Impact Profile. that incorporate the patient's self-assessment. These instruments have proved useful in enabling more consistent comparisons of disease and treatment impact across conditions, and in enabling the treatment-specific impacts of care on a patient life to be detectable even when the patient has multiple health conditions (see chapter 3). The National Eye Institute, for example, uses one or both instruments in at least three of its clinical trials, and several trials of AIDS treatments use a variation of the SF-36 adapted for that particular condition (846 ).

Most NIH trials that incorporate patient functioning or quality-of-life as an outcome measure. however. apparently use disease-spccif:c instru-

## BOX 4-2: The Elusive Prostate Treatment Trial

The outstanding example of a comparative effectiveness trial that did not happen was the result of efforts to investigate alternative therapies for benign prostatic hyperplasia (BPH), the subject of one of the first four Patient Outcomes Research Teams (PORTS) funded by the Agency for Health Care Policy and Research (AHCPR).

Research results from the BPH PORT documented enormous variation in the rates at which physicians chose to treat this condition with early surgery (as opposed to "watchful waiting, " then surgery only if symptoms worsened) (918) Results from database analyses also raised questions about the relative effectiveness of new transurethral surgical procedures compared with traditional open surgery (920). Although early suggestions that the transurethral procedure was actually less safe were probably unwarranted (1 40), the research nonetheless raised significant questions about the effectiveness of alternative management strategies that prompted the urological community to consider a randomized study of alternative treatments for the first time.

In fact, the American Urological Association (AUA) proposed such a clinical trial and applied to AHCPR for trial support (41). The AUA also conducted a pilot study of 400 patients to demonstrate the feasibility of the idea (913).

AHCPR concluded that the study initially proposed was too expensive to be feasibly funded out of the agency's small budget. The AUA then submitted a second scaled-down proposal, but it was deemed by reviewers unlikely to be large enough to answer the questions being investigated (41), The National Institutes of Health (NIH), on its part, was apparently uninterested in funding a study that involved only existing treatments and offered little opportunity for new insights into the underlying biological mechanisms of the disease

Paradoxically, other studies of treatments for BPH are taking place that in their way highlight the current inadequacies of effectiveness research Both NIH and the Department of Veterans Affairs (VA) are conducting randomized clinical trials testing finasteride, a newly approved drug that is

ments. In fact, NIH staff note the contributions of NIH-sponsored research in developing instruments for such disease areas as cancer and rheumatology. They also note the current efforts to develop a standard vision function questionnaire suitable for patient self-assessment of the broad spectrum of vision function deficits.

Although the diversity of measures being studied and applied has advantages, some agreement on common measures is desirable for the sake of making cross-study comparisons, at least within the same disease. The efforts of the National Eye Institute in supporting development of an instrument useful for studies of eye disease, for example, are aimed at this goal (852).

The fact that a majority of comparative clinical studies still apparently do not include data collection on patient functioning and quality of life, whether through generic or disease-specific instruments, suggests that opportunities are being lost to provide important information for future patient decisionmaking. Equally troubling is that efforts to fill this need do not seem to be proceeding easily. AHCPR and NIH both clearly have much to contribute on the topic, yet cross-fertilization across the agencies in this area is more

**BOX 4-2 continued: The Elusive Prostate Treatment Trial**

believed to reduce symptoms in patients with this condition. The NIH trial, currently in the pilot phase, Involves 150 men at six medical centers The proposed full trial will be larger and more extensive, lasting six years (859a,879) In contrast, the VA trial will involve 1,200 men at 30 VA medical centers for only a year (564a) Trial designs are somewhat different as well The NIH trial is comparing finasteride against an alternative drug and a placebo and will revolve a number of tests and measurements aimed at better understanding the underlying disease The VA trial is likewlse testing the drug against both an alternate drug and a placebo but with a much simpler protocol and fewer clinical measurements (132)

There are two main differences between these funded studies and the unfunded one that was proposed to answer some of the questions raised by the PORT The first is that the funded studies involve a new technology, the drug flnasteride NIH's interest in funding a trial is piqued much more by new than by existing technologies, particularly when the trial offers possibilities for additional biochemical research as well Second, the funded studies involved a drug rather than a procedure Drugs unlike procedures, must be approved by the Food and Drug Administration (FDA), and manufacturers are accustomed to the routine of clinical trials Drugs also have identifiable "owners" who can profit from the results and thus sometimes may be willing to help support a study, in fact, the VA study is receiving support from the manufacturers of both drugs being tested (564a)

Both trials have worthwhile goals, and some duplication in research can add to the validity of the overall findings Still, in an area of research in which the gaps are so great, and the resources being made available to fund clinical trials on existing therapies so limited, it is ironic that the federal government is funding two simultaneous studies of a single therapy for benign prostatic hyperplasia, when another study of the same disease that was clearly needed has been unable to find funding

SOURCE Off Ice of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

notable for the exceptions than the norm. (In one exception. for example, AHCPR and NHLBI are collaborating to incorporate a quality-of-life measure into the ALLHAT[3] clinical trial of hypertension therapies (821).) Many institutes seem to have relatively little interest in the methodological work done at AHCPR; and where there is interest, AHCPR seems to perceive it as interest in that agency's resources rathcr than real interest in intellectual collaboration. [4]

### *Enhancing Know/edge of Ongoing Trials*

Making the most use of ongoing clinical trials requires knowing that those trials are happening. and something about their characteristics. It has been suggested that one way to do this is to create a register of all ongoing clinical trials ( 166). Potential purposes of such a register are:

1. to foster more efficient research spending. by promoting collaboration among investigators

[3] AI, I. HAT is the acrony m for the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial.

[4] A recently published Italian study of the comparative effectiveness of different followup regimens for breast cancer patients underscores the feasibility and potential benefits of incorporating quality-of-life measures in effectiveness trials (280).

considering similar trials and preventing unnecessary duplication of research,
2. to enable better methodological research about the way that trials are undertaken and used (e.g., studies into publication bias of research results);
3. to recruit patients and providers into clinical trials more effectively and efficiently; and
4. to enhance scientific reviews of the literature, including meta-analyses (166).

The third of these reasons addresses the needs of effectiveness research in an especially direct manner. If broad groups of patients from across geographic areas and care settings are to be included in trials so that trial results areas generally applicable as possible, patients and their clinicians must know about trials. And, if large, multisite trials are to be completed in time for their results to be useful, patients must be enrolled as quickly as possible.

A number of registries of ongoing clinical trials in particular topic areas do exist. The AIDS and cancer communities have been particularly active in supporting registries so that patients and clinicians can learn about ongoing trials for which they might qualify. The PDQ database of the National Cancer Institute, for example, contains information on ongoing and completed clinical trials of cancer therapies in the United States. Information on AIDS treatment trials is available through MEDLINE, and numerous regional AIDS information services include additional detail on trials in their areas (167,418,757).

The AIDS database of ongoing clinical trials is unique because it relies on a special statutory except ion for information to be released by the Food and Drug Administration (FDA). Until 1988, FDA was prohibited from releasing information on ongoing clinical trials funded by private industry, information that industry generally considers to be confidential (418). This prohibition was lifted for information on AIDS-related trials only, due to the urgency of research on this disease. Thus, both NIH and FDA contribute information regarding ongoing trials, so that both publicly and privately sponsored clinical trials are included in the database. Information on private trials conducted under FDA auspices is summarized at FDA to protect as much confidential information as possible (191). In contrast, the PDQ database includes all NCI-sponsored trials, but it includes information on other trials only if that information is volunteered by the sponsoring organization (757).

The only cross-topic registries of ongoing controlled clinical trials in the United States are the clinical trials databases held by VA and NIH, respectively, to keep a comprehensive list of the clinical trials they sponsor. The NIH database is of special interest, because NIH is such a prominent sponsor of clinical trials, and because these trials are less 1 inked to a particular demographic population (i.e., veterans).

NIH maintained an inventory of its clinical trials from 1974 until 1979, when it discontinued the inventory for budgetary reasons (864). The inventory was re-established in 1985, through the Office of Medical Applications of Research (OMAR). Its road, however, has been rocky. Data collection was onerous; rules were changed in 1988 to require Institutes to report only data on controlled clinical trials (data on uncontrolled trials became optional). Data on mechanisms and sources of financial support has been particularly difficult to collect consistently (235). Data on cancer trials does not correspond precisely to data on trials from other Institutes, due to NCI's own internal trial database (864).[5]

Recent legislation requiring NIH to compile a cross-disease registry of clinical trials that involve women has helped stimulate interest in assembling a comprehensive database of ongoing clini-

[5] For all of these frustrations and limitations, summary data from the database, available for 1989, are interesting. In that year, NIH supported 440 controlled c1 inical trials, at an average annual cost per trial of just under $800,000 (864). NCI trials were excluded from the calculation of average annual cost per trial due to data inconsistencies.

cal trials (235). One barrier to any comprehensive registry is the lack of incentive for privately funded trials to be included: manufacturers consider this information confidential. FDA regulations protect this confidentiality, in order to protect manufacturers' financial incentives to develop new products. Short of withdrawing this protection, any comprehensive database must rely on the voluntary participation of private sponsors.

The other major barrier to a comprehensive clinical trials database is cost. One suggested solution is to use electronic technology to link existing trial databases. rather than initiating new ones, although this suggestion suffers the constraint of being limited in scope to the topics of existing (or new) registries (865).

Although there is clearly some interest in comprehensive, or linked, registries of ongoing trials, there is as yet no consensus about what form such an effort should take if it happens, or even about the extent of information such a database should contain. As NIH staff point out based on their experiences trying to maintain an NIH-wide trial database, collecting more detail on each trial would make the database more useful to researchers, trial participants, and policy makers alike, but greater detail comes at the cost of greater difficulty obtaining complete, accurate. and consistent data from the Institutes themselves (235).

### *Improving the Clinical Research Infrastructure*

An important component of effectiveness research is the effort to make study results relevant to ordinary practice and the population at large. To do so, studies must address issues that arise in everyday care. and they must include an array of patients and providers representative of the overall population. For many questions, such as those in the area of primary care. undertaking comparative effectiveness trials can require large numbers of patients and physicians who are not currently affiliated with research institutions. The financial and administrative barriers to setting up such trials are substantial, and a major reason why these trials are not more common (box 4-3 ).

Furthermore, the barriers to large-scale, community-based trials must be overcome anew for each new trial proposed. NHLBI, for instance. is investing considerable resources in establishing a research network with as many as 300 practice sites for its ALLHAT trial of antihypertensive and cholesterol-lowering therapies (846). Once the trial is over, however, the network may well dissolve.

Conducting broad community-based trials would be substantially more streamlined if a network of providers already existed who had previously agreed to participate in research of interest to them and their patients. Establishing, maintaining, and supporting such networks is one way that the federal government could enhance the efficiency of comparative effectiveness trials, the generalizability of their results, and researchers' ability to carry them out.

To increase both provider and patient participation in clinical trials, trial enrollment and data collection requirements may need to be simpler than they are in many current trials. Thus, those designing and funding clinical trials may need to pay more attention to the techniques of large, simple trials described in the previous chapter. In addition, however, researchers and sponsors must find ways to recruit, train. and support a much broader set of very busy practicing clinicians.

Some of the best known examples of standing research networks are the infrastructures created for the various large, simple trials of therapies for heart disease. The GISSI and ISIS trials, described in chapter 3, both created an infrastructure of participating hospitals in their first respective studies that could be used on future trials as well; the fifth trial using the ISIS network is now underway. The important point about these networks is that they include many centers that are not teaching institutions and otherwise might not participate in detailed clinical trials. The GISSI network is an interesting model because it is so comprehensive: most of the coronary care units in Italy have participated in the GISSI trials.

Several U.S. examples of community-based medical research networks exist as well. A num-

**BOX 4-3: The Need for Community-Based Research Networks: A Hypothetical Example**

The administrative barriers to conducting large-scale, community based research are substantial. Establishing and carrying out such a study usually requires a major investment in recruiting providers and patients to participate. The investment is especially a great barrier for comparative testing of technologies already in use, since there are few eager sponsors for experiments involving existing interventions

As an example of the administrative difficulties a potential trial might face, imagine a researcher wishing to conduct a clinical trial of the comparative effectiveness of two common medications (inhaled cromolyn vs Inhaled steroid) in enabling the maintenance of normal activities in children with mild asthma Most of these children would be managed by primary care physicians, and many would never have even been hospitalized for their condition. Furthermore, effectiveness could well vary according to characteristics of children (e. g., cromolyn might be presumed to require more doses per day to be equally effective, and compliance with this stiffer regimen might differ according to a child's age)

Thus, the trial would have to recruit a large number of children covering a wide range of ages and other characteristics The researcher would need to identify these children, recruit their physicians, train these physicians, and have funding sufficient to give them the support they need to follow the study protocol and collect data for the trial. Simply getting the trial underway and convincing physicians to participate in the study would require a major investment of time and resources

Even when administrative barriers to such a trial are overcome, financial support may not be forthcoming. The American Urological Association, for example, tentatively established a network of physicians willing to participate in an ongoing series of trials of therapies for prostate disease (1 32) The network has never become fully operational, however, because the initial trial was never funded (see also box 4-2)

SOURCE Off Ice of Technology Assessment, 1994

ber of small practice networks exist that are loosely organized but enable clinicians and researchers to connect as needed to address particular research questions (122). The VA Cooperative Studies program is a standing multisite program that routinely involves practicing clinicians in clinical trials (523). Three additional examples illustrate in more detail the potential and experience so far with practice networks.

The *Community Clinical Oncology Program (CCOP),* sponsored by the National Cancer Institute (NCI), supports patients and physicians in community hospitals who wish to participate in cancer trials. NCI provides funding to cover administrative and data collection costs, without which community hospitals might not be able to participate in trials. The trials themselves are coordinated by NCI-supported teaching and research hospitals (260). About 50 CCOPs, representing about 300 community hospitals, receive funding from NCI to support their participation in cancer trials through this network. CCOP patients represent roughly one-third of all patients enrolled in NCI trials (260).

The *Ambulatory Sentinel Practice Network (ASPN),* a private effort supported in part by the American Academy of Family Practice, is another longstanding U.S. community research network. Established in 1982, its purpose is "to increase and refine the primary care knowledge base by

studying the problems that occur in primary care" (11 ). It includes 72 participating medical practices (including over 300 practitioners) in the United States and Canada. An overwhelming majority of the participants are family practice physicians (12).

Because its members are largely community-based primary care physicians, ASPN's data collection has been very simple: basic demographic data on patients seen in the practice, with data collection on the study question through a weekly mailed card. Many study questions originate with the practitioners themselves, and most are descriptive studies. Funding for individual studies is sought from whatever sources are available: sponsors have included such federal agencies as the Centers for Disease Control and Prevention and NHLBI. Examples of recent and ongoing studies include depression in primary care; management of carpal tunnel syndrome; acute low back pain; and the effect of digitalis on mortality ( 12). Research is administered through a central headquarters in Denver, Colorado.

A possible concern of this and other research networks is that because the participants are self-selected, their patient populations may not be representative of patients overall. ASPN researchers addressed this concern by comparing detailed characteristics of patients and visits to ASPN practices with the characteristics reported on a national survey of ambulatory care (the National Ambulatory Medical Care Survey) (297). They found considerable similarity in visit characteristics (e.g., patient diagnoses) but some differences in patient demographics.

The *Vermont Trials Network is* a newer private effort, an innovative network of hospital neonatal intensive care units established to perform collaborative clinical research in neonatology and integrate research into daily practice (383). As of February 1994, there were 111 neonatal centers participating in the network, many of which had no affiliations with universities (729). The great majority of these are centers in U.S. hospitals, but recently hospitals in Australia, Germany, Japan, and other countries have also expressed an interest in participating (729).

The centers collect basic data on the medical and demographic characteristics of infants. They also collect information on the prevalence of some conditions and on the use of particular technologies and services (e.g.. the use of ventilators and surfactant). These data are intended to provide information for planning clinical trials and to pm-nit centers to compare their outcomes with each other as an aid in quality management (354). The database and trials facilitation service are administered through a central office in Vermont, which operates with temporary grant funding from a private foundation.

The first clinical trial to be implemented in the network centers, which began in January 1992, was a randomized comparison of two commercially available surfactants (preventive treatment for lung disease in premature infants). Both surfactants have been proven effective in previous trials, but direct comparisons of the two drugs are not available (354). The participating researchers hope to find an answer of practical importance to community neonatologists and to be able to compare the costs and results of this trial to those of a smaller, NIH-funded trial on the same topic being carried out only at university centers (354).

These three examples differ considerably in their sophistication. sources of funding. and size. They range from research by office-based family practitioners to large clinical trials in neonatal care units. What all have in common is that they involve an underlying structure through which non-academic as well as academic health care providers can participate in clinical research of interest to them and their patients. Indeed, in the case of the ASPN network, the providers themselves suggest some of the research questions.

None of these examples are of "firms" research infrastructures, which may require more intensive effort and investment on the part of the health care institution. The emergence and growth of managed care providers and the interest in methods for continuous quality improvement. however, might

make firms structures an attractive form of research for many institutions, particularly if they received some startup financial support. The VA is exploring the establishment of a research structure of this type (930).

An aspect of comparative effectiveness trials largely uncommented on, in either the literature or the health policy debate, is the relationship between effectiveness trials conducted within a committed infrastructure and the goals of continuous quality improvement, a topic that is very much the subject of current discussion. This relationship is particularly marked in the GISSI large, simple trials, which included most of the coronary care units in Italy. As the trials were completed, units could incorporate the findings, and new trials begun to achieve the next level of quality improvement. Questions of generalizability of findings were almost irrelevant, since most units and patients participated. Firms trials have accomplished this objective on an institute-specific basis; as an intervention proved effective, it was adopted by the other firms in the institutions and became the new level against which future improvements would be measured.

## The Comparative Evaluation of New Technologies

A major contributor to the current state of ignorance about what works best, and under what circumstances, in health care is the fact that many—probably most-new medical technologies need not undergo rigorous review of their effectiveness before being adopted by practitioners and patients. Furthermore, of those that are reviewed for their effectiveness, most need not prove that they are actually more effective than other alternative technologies already on the market.

There are three avenues through which new technologies can be identified and enrolled in comparative evaluations:

- **Manufacturers.** Those producing new technologies could be encouraged to identify them and conduct comparative assessments directly. This avenue could take the form of increased regulatory oversight, such as a broad extension of current FDA requirements for new drugs; or it could take the form of inducements (e.g., favored regulatory treatment for manufacturers willing to sponsor comparative postmarketing studies).
- **Payers.** Health insurers, including government payers, could offer insurance coverage for new technologies only if they had met explicit standards of evaluation and effectiveness.
- **Government.** Some researchers have suggested that the most efficient way to increase the number of direct comparative studies on new as well as existing technologies is for the federal government to conductor sponsor such studies directly (624,625 ).

These avenues are not mutually exclusive; all three could be pursued simultaneously.

An underlying question implicit in choosing among these options is who should be paying for the evaluation of new technologies. Manufacturer-sponsored evaluations could come about either through regulatory incentives or pressure by payers. Alternatively, payers could withhold coverage from unevaluated new technologies but could also help fund their evaluations, by paying for some of the costs of the studies. Government-sponsored evaluation would clearly increase the proportion of studies of new technologies funded by taxpayers generally.

Three issues are especially prominent in considering how to enhance the number and quality of comparative evaluations of new with existing technologies. The first, especially important in strategies that depend on manufacturers to conduct evaluations, is the role the FDA plays in the evaluation of new technologies. The second issue, associated with payer-dependent strategies, is the role of health insurers in paying for new and experimental technologies. Both of those issues are discussed in this section. The third issue is the potential role of different federal agencies in conducting or supporting evaluations. This issue extends to current effectiveness research efforts

comparing existing technologies as well, and it is discussed in the final section of this chapter.

### *Role of the Food and Drug Administration*

The charge of the FDA is to ensure that new drugs and medical devices are safe and efficacious--i.e., that the medical benefits outweigh the medical risks—before they are marketed to the public. Its regulatory authority extends not only to whether a product can be put on the market, but what claims the manufacturer can make about that product. In reviewing evidence about the efficacy of a product. FDA gives strong weight to evidence from randomized clinical trials as the most valid basis for making efficacy claims.

FDA's authority over medical devices is slightly different from its authority over drugs. All drugs that involve new chemical formulations must show proof of efficacy, with a stringent level of evidence to provide that proof. Most often they are compared in randomized trials with placebos, although new drugs in certain categories, such as new antibiotic and anticancer drugs, are commonly tested against accepted existing drugs instead (748). In contrast, new medical devices are categorized by FDA staff into one of three classes, according to the types and controllability of risk associated with the device in its intended use, with each class subject to a different standard. Class I and II devices considered to involve only low or moderate risk-e. g., new wheelchairs—must be registered with FDA, and their producers must conform to good manufacturing standards. Class III devices such as x-ray machines also must meet performance standards. In addition, however, Class III devices—generally those posing a potentially higher risk to patient health—must meet standards similar to those for new drugs. [h] Class 111 devices account for roughly 10 percent of medical devices (922).

Over time, FDA policies have changed somewhat in the kinds of outcomes considered the most relevant for regulatory decisions. For medical devices, the agency has historically placed a strong emphasis on what FDA terms "functional utility:" i.e., whether the device does what the manufacturer claims it does (e.g., remove plaque in blood vessels). In 1990, an internal FDA policy guideline established 'clinical utility''—the ability of the device to produce a desirable treatment outcome—as a preferable standard. Under this standard, for example, home uterine monitoring devices would have to prove not only that they could detect uterine contractions, but that clinical outcomes (e.g., the number of premature births) were improved (922).

Trends in the standards for evaluating new drugs have some differences from those for devices. In some areas, for example, the trend has been to emphasize clinical endpoints that can be measured quickly. In particular. the urgency of the need to identify drugs that might be efficacious in treating AIDS has led to greater use of "surrogate endpoints" in the approval of anti-AIDS drugs for marketing (e.g., endpoints such as showing a difference in the rate of certain biochemical markers that indicate the progression of disease). The use of surrogate endpoints has its own well-known hazards; in a recent example, a drug approved for marketing by FDA on the basis of improvements in surrogate endpoints could not be shown, in a longer European trial, to have any effect on total mortality from AIDS ( 142). FDA staff cite this example as a reason to conduct post-marketing studies of such drugs, so that effects on ultimate endpoints can be measured as well (839).

In other areas, however, there are examples of a greater attention to ultimate outcomes (e.g., mortality) as a factor in FDA decisionmaking. In the clearest example, quinidine--a drug originally approved for marketing on the grounds that it was shown to be efficacious in reducing atria] fibrillation (irregular heartbeats)—was later required to be relabeled or withdrawn from the

[6] Class III devices considered by FDA to be "substantially equivalent" to a device already on the market in 1976. when the regulatory authority over medical devices was added, are not immediately required to meet these standards but can be required to do so in the future (784a).

market after a clinical trial showed that this drug actually increased, rather than decreased, mortality rates in some groups of patients (839).

The randomized clinical trial continues to be the gold standard for the assessment of a product efficacy, but some of the more recent innovations in conducting and analyzing clinical trials are occasionally finding their way into FDA decisionmaking. Large, simple trials, for example, have not been used as a basis for approving a drug, but they have been used to support approved changes in a drug's label or advertising (839). Similarly, the results of a meta-analysis have been used as the basis for insisting that a drug be relabeled, after the meta-analysis showed treatment groups to have a higher overall mortality (839).

An interesting example of "effectiveness" trials required by FDA involves the transition of a drug from prescription-only to over-the-counter availability. Manufacturers interested in marketing drugs for nonprescription uses must conduct "usage trials." In a typical usage trial, several thousand patients are given the medication, with its proposed labeling and instructions for use, and are monitored to determine whether the drug is safe and effective as actually used by these patients (327).

Thus, FDA plays three strong roles in the comparative evaluation of new medical products.

- First, it requires that the underlying efficacy of all new drugs, and some new devices, is established—i.e., that the product works under at least some conditions. Efficacy sometimes involves direct comparisons with existing drugs, as in the case of antibiotics, but even direct comparisons often do not provide broad information on comparative effects in ordinary practice.
- Second, although much of FDA's role in drug and device approval focuses on approving new products for marketing, the agency also plays a role in the post-marketing monitoring of the effects of products in general use, and it plays a strong role in the effectiveness claims that manufacturers can make when advertising their products.
- Third, FDA establishes acceptable levels of evidence for showing that a product works. Hence, FDA's greater emphasis on ultimate health outcomes, and on results from randomized trials, have trickle-down effects on health research. Standards have generally required that a trial show a very strong ability to reject a hypothesis that the new treatment made no difference, an issue that has potential repercussions for FDA oversight of later comparative effectiveness and cost-effectiveness claims.

### *Issues in Insurance Coverage of Newly Introduced Technologies*

The role of payers in covering (or withholding coverage from) new or experimental technologies is an issue that has been growing in prominence. Its importance to government policy makers contemplating changes to the health care system is demonstrated by proposals to include experimental services as a health insurance benefit under certain conditions. The State of Maryland, for example, is devising a basic benefit package, which insurers who market health insurance to small employers must offer, that includes coverage for technologies offered as part of authorized clinical trials (409). The health reform proposal of the Clinton Administration included a provision that would have required coverage of "routine care" associated with experimental therapies and permit coverage of the therapies themselves if they met certain conditions (S 1757).

Historically, insurers have relied on the term "medically necessary" to broad] y describe the services covered by their health policies and "experimental" to define at least some of the services beyond the boundary of health care coverage. Since the 1980s, the definition of these terms has been an increasingly contentious issue. Experimental services are particularly controversial because they often involve potentially life-saving treatments for desperate y ill patients who are personally willing to take the risk that the service may prove to be unsafe or ineffective. Today, the interpretations of "medically necessary" and "experimental" are hotly contested among insurers,

researchers, physicians, manufacturers of drugs and devices. and patients and are often mediated (albeit inconsistently) in the courts.

A typical insurance contract defines a service or supply to be "medically necessary," and therefore covered, if: (a) it is ordered by a doctor; (b) it is commonly and customarily recognized throughout the doctor's profession as appropriate in the treatment of the sickness or injury; and (c) it is "neither educational nor experimental in nature nor provided primarily for research purposes" *(322 ).7* The point at which a new treatment moves from the investigational or experimental stage and into the realm of '-state of the art" medically necessary treatment is not at all clear (25).

There are no data that systematically document those services commonly excluded by insurers because of their experimental nature. Autologous bone mm-row transplant with high-dosage chemotherapy (ABMT/HDC) for breast cancer is perhaps the most widely contested and well-known experimental treatment (box 4-4). Other examples of technologies typically excluded from coverage on the grounds that they are currently experimental, or covered only case-by-case, include growth hormone for children with short stature, pancreas transplants, and home uterine monitoring for the prevention of premature births ( 178).

Despite explicit contract language to the contrary, it appears that some insurers sometimes allow coverage of certain experimental treatments on a case-by-case basis. For example, five major carriers reported in a recent telephone survey that, given certain criteria and conditions. they would pay for a number of "experimental" treatments including ABMT/HDC. pancreatic transplant, growth hormone for short-stature children, home uterine monitoring, and radial keratotomy ( 178). Researchers conducting a clinical trial comparing ABMT/HDC with conventional treatment point out that coverage decisions across and even within insurance companies for this therapy are inconsistent (601 ).

Recently, a few insurers have taken an unprecedented step into the controversy surrounding: coverage of newly introduced technologies. They have agreed to pay for ABMT/HDC for insured patients who are enrolled in an NCI-approved randomized controlled clinical trial to compare ABMT/HDC to standard treatment for breast cancer (404) (box 4-4). An important component of these trials is that they are randomized: thus, many of the patients enrolled in the trial will not receive the experimental therapy. For this and other reasons, patient accrual to the trials has been disappointing to researchers ( 135).

The issue of coverage for experimental technologies has begun to receive attention from insurers at a national level as well. The Health Insurance Association of America ( H IAA), for example. has endorsed a policy to encourage their membership to pay for the patient care costs related to NIH-sponsored and certain other officially endorsed randomized clinical trials. The National Association of Insurance Commissioners has established a working group on the topic. whose goals include drafting a model regulation or statute to address off-label use of prescription drugs, and researching the impact of experimental treatment exclusions. Several States (e.g., Washington, Florida, New Hampshire) have already passed insurance regulations related to experimental treatment (552).

Despite the increasing interest and movement to change the link between insurance coverage and the experimental status of a technology', proposals to address the connection between coverage and the degree to which a technology has been proven effective face a number of competing interests and concerns:

7 An interesting facet of the definitions that exclude experimental technologies from insurance coverage is demonstrated by contract language for insurance contract exclusions developed by Towers, Perrin, Inc. for its c 1 ients use in their health benefit plans (178). I n this contract language, the fact that a technology is the subject of a controlled clinical trial merits the label of experimental. This categorization might present a problem for randomized c1 inical trials comparing two technologies already in common use.

## BOX 4-4: Autologous Bone Marrow Transplants in the Treatment of Breast Cancer

Conventional therapy for women with advanced breast cancer consists of mastectomy followed by radiation, chemotherapy, or both, But conventional therapy frequently fails, and approximately 46,000 women die of breast cancer each year (777),

A limitation of conventional chemotherapy is that it cannot be administered at high dosages without killing the patient's own bone marrow as well as the cancer cells. Autologous bone marrow transplant with high-dose chemotherapy (ABMT/HDC) is a technique aimed at enabling higher doses of chemotherapy to be given. In this procedure, the patient's bone marrow is removed before chemotherapy is administered, and then reinfused after the chemotherapy regimen is complete,

ABMT/HDC has engendered great enthusiasm in the medical world (670) and is now being tried for other solid tumor cancers as well (e.g., testicular and colon cancers). Still, the efficacy of ABMT/HDC over standard treatment for advanced breast cancer has not yet been definitively demonstrated (850), and one assessment of the technique based on past studies expresses skepticism that it wiII prove effective in this population (195).

Most payers view ABMT/HDC for metastatic breast cancer as experimental, Recently, however, some major insurers and HMOs (including Metropolitan Life, Prudential, CIGNA, Travelers, U.S. Healthcare, Kaiser, and some Blue Cross Blue Shield plans) have begun to cover ABMT/HDC for breast cancer under certain conditions (66,1 78,404). It is not clear that these private health insurers have taken this step because they now accept ABMT/HDC as nonexperimental or state-of-the-art therapy. Rather, there is much anecdotal evidence to suggest that they are motivated by the desire to avoid legal action, expense, and negative publicity. Efforts by insurers to refuse reimbursement for ABMT/HDC for breast cancer have been widely contested in the courts; in one recent and well-publicized case, a jury awarded $89 million to the patient of a California HMO that had refused to cover the procedure (135).

Still, insurers' response to the pressure to pay for ABMT/HDC have varied widely, not only among carriers but within companies as well. For example, in an effort to develop the clinical data necessary to assess ABMT/HDC's efficacy, U.S. Healthcare[1] and 17 Blue Cross Blue Shield plans[2] are currently supporting several National Cancer Institute (NCI) randomized controlled clinical trials to compare ABMT/HDC to standard treatment for breast cancer (66,404). The trials are continuing to accrue patients (with varying rates of success) and final results are not expected for at least three years (135). Other insurers, including Metropolitan, CIGNA, Prudential, Travelers, and Aetna, are now reimbursing for ABMT/HDC on at least a case-by-case basis, but apparently do not require that patients participate in the NC I trials.

---

[1] U S Healthcare is a 16 million member HMO based In Philadelphia, PA

[2] These 17 plans include approximately half of the nation's Blue Cross and Blue Shield's membership (282)

SOURCE Office of Technology Assessment, 1994, based on sources as shown Full citations at the end of the report

- Manufacturers generally support coverage for technologies that are still at the experimental stage as a way to minimize the time lag between a product's development and its availability to patients (461 ). At the same time, a proposal that linked payment to rigorous proof of efficacy would probably meet producer resistance, since many existing technologies cannot meet this standard (and many new ones need not at present). Whether a standard of proof as rigorous as the RCT is even necessary for all technologies is very much a matter of debate.
- Patients and providers likewise generally support coverage for experimental technologies, since it would increase the treatment options financially available to them.
- Insurers, and those who pay the insurance premiums, tend not to support coverage of investigational interventions, on the grounds that coverage would increase costs without any assurance that the interventions would be either safe or effective for those who would receive them.
- Some observers also express concern that opening insurance coverage to investigational therapies could lead to a worsening of the problem of poorly conducted studies, unless strict controls and monitoring of the investigational protocols were also in place (581).

## ISSUES IN FEDERAL FUNDING AND SUPPORT

### The Roles of the Federal Agencies

The self-perceived roles and goals of agencies have a strong influence over the part each plays in the current debates over how to improve the effectiveness, quality, and costs of health care. They also explain a great deal about where and why duplication or gaps in effectiveness research appear among agency activities.

The federal organizations that currently sponsor effectiveness research (and other evaluative activities) do so for three reasons. The first is to provide information to the public and to private insurers and providers, in order to improve the private sector's ability to deliver effective care. The second is to support the government own health care financing and delivery programs. such as Medicare, the veterans' health system, and the myriad preventive and other public health programs. The third reason is to provide information that can enhance public policy decisionmaking generally, for purposes ranging from distributing research resources to helping Congress decide whether to establish new Medicare benefits.

Most of the federal organizations involved are sprinkled throughout the U.S. Department of Health and Human Services (DHHS). They include the Agency for Health Care Policy and Research (AHCPR), the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), the Health Care Financing Administration (HCFA), and a small office under the Assistant Secretary for Health, the Office of Disease Prevention and Health Promotion (ODPHP) (figure 4-1, see p, 105). In addition, however, the Department of Veterans Affairs (VA), which operates a health care system for veterans of the U.S. armed services, has its own entirely autonomous research arm to investigate health care services and technologies.[8]

These organizations all are concerned in some way with health care research and delivery, but they have greatly differing purposes and orientations. These differing purposes affect their approaches to identifying effective and cost-effective medical services and technologies, their methods for assessing technologies for clinical and public policy purposes, and the degree of their activity in these areas.

[8]To identify relevant activities currently conducted by these agencies, OTA asked administrators in each organization to provide information on those studies and activities they considered relevant. Their responses form the basis for discussions of activities presented in this report, The Department of Defense also conducts medical research, but its activities were not investigated in detail in this report.

**BOX 4-5: The Components of the Agency for Health Care Policy and Research**

1. The **Center for General Health Services *Extramural* Research** carries on much of the legacy of general research into the interactions of the health delivery system inherited from the National Center for Health Services Research (NCHSR) The center supports research on such topics as health care costs and financing, and Improving the delivery of health care services to special populations.
2. The **Center for General Health Services *Intramural* Research** is another NCHSR legacy. It performs in-house general health services research, drawing heavily on data from federal health care databases.
3. The **Center for Medical Effectiveness Research** is the focal point for the federal government's investment in effectiveness research. This center supports the Patient Outcomes Research Teams (PORTS), as well as many other extramural research projects on medical practice and outcomes variation, the effectiveness of particular medical interventions, and the refinement of some of the tools of effectiveness research.
4. The **Office of Science and Data Development** oversees activities related to the enhancement of databases and the Implications of advances in medical information systems. Its activities support the effectiveness research infrastructure—e.g., by supporting efforts to link large databases together, and sponsoring research to stimulate the development computer-based patient records
5. The **Office of the Forum for Quality and Effectiveness in Health Care** is responsible for the development of clinical practice guidelines. It organizes the guidelines panels and provides them with staff support and supplemental contracted expertise.
6. The **Office of Health Technology Assessment** is another direct holdover from the old NCHSR, although its responsibilities have expanded somewhat. It conducts In-house assessments of individual medical technologies for the Medicare and CHAMPUS programs.
7. The **Center for Research Dissemination and Liaison** has the primary responsibility for disseminating clinical practice guidelines developed by the Forum to clinicians, patients, and other Interested parties It also operates the Users Liaison program, which runs informational conferences and provides technical assistance to State personnel and other consumers of AHCPR's work

SOURCE Off Ice of Technology Assessment 1994, based on documents provided by the U S Department of Health and Human Services, Public Health Service Agency fc. Health Care Policy and Research, Rockville. MD 1993

## Agency for Health Care Policy and Research

Implicit in Congress' creation of AHCPR in 1989 was a statement that the federal government should actively support and promote effectiveness research and health technology assessment. AHCPR is primarily a research-sponsoring organization. with an agenda that tends to focus on health services and current therapies, the legacy of its inheritance of the National Center for Health Services Research (NCHSR), its predecessor agency. It contains seven centers and offices (BOX 4-5). Three of these are direct holdovers from NCHSR. The remaining four were newly created specifically to carry out AHCPR's new mission.

In conformance with its small budget, its health services research orientation, and its legislative mandate, AHCPR's investment in effectiveness research has leaned heavily towards the development of effectiveness research tools (e.g., developing databases and health status measurement instruments) and descriptive research on the outcomes associated with particular technologies

and patterns of care. Many of these activities take place through the Patient Outcomes Research Teams (PORTS), the central research program of the federal government's effectiveness research initiative As described in chapter 3, AHCPR supports very few controlled trials of clinical interventions.

AHCPR's budget grew from $97 million in fiscal year 1990 to $128 million in fiscal year 1993, All of that increase went to effectiveness research and guideline development efforts, whose funding grew from $37 million $73 million during the same period (J. Clinton, at AHSR, June 1993). The budget is broken into three activities:

- Program support. This component receives the smallest portion of the budget—$2.5 million in 1993, or about 2 percent of the total
- Research on health costs, quality, and access. This component is the continuation of the health services research efforts previously carried out through NCHSR and amounted to $53.1 million in 1993, or 41 percent of the total. It supports both intramural and extramural general health services research and supports the National Medical Expenditures Survey, a major source of medical cost data.
- Medical **Treatment Effectiveness Program** (MEDTEP). This funding line accounted for $73.0 million of AHCPR's 1993 budget, or 57 percent of the total. It supports not only the acti - vities sponsored by the Center for Medical Effectiveness Research, which funds extramural effectivencss research, but also the guidelines activities of the Forum, the resource development activities of the Office of Science and Data Development, and the activities of the Center for Research Dissemination and Liaison.[9]

Like most agencies, the great bulk of AHCPR's funding (85 pcrccnt,or$109 million in fiscal year 1993) clinics from federal gcncral revenues funding. AHCPR's authorizing legislation also permits substantial transfers from the Medicare Trust Fund for medical effectiveness research and guidelines development. In fiscal year 1993. $103.6 million was authorized from this source, but only $5.8 million was appropriated. making up 4 percent of the *agency's* budget (814). In addition, AHCPR is authorized to draw funds from the Public Health Service Evaluation Set Aside ("One Percent Funds"). These funds account for a significant proportion of the agency's total budget ($1 3.2 million in fiscal year 1993. or 11 percent of the budget), but they are earmarked to fund the National Medical Expenditures Survey and cannot be used for other purposes under current law.

### *The National Institutes of Health*

The National Institutes of Health (NIH), with a budget of approximately $10 billion in fiscal year 1993, is the primary sponsor of biomedical research in the United States (844). From its origins in a public health service research laboratory (box 4-6). NIH has come to comprise 24 relatively independent research institutes. NIH coordinates an extensive intramural research agenda as well as funding extramural research conducted at 1,700 institutions nationwide.

Most NIH institutes conduct a great amount of basic "'bench" research as well as some applied clinical research, primarily on developing new therapies. NIH spent $864 million, or just under 10 percent of its budget ($8.4 billion), on clinical studies in 1992 (844). Although in recent years a significant proportion of NIH funding has been "earmarked" (e.g., for AIDS or women health research), its agenda is still largely investigator-driven and heavily influenced by the makeup of its "study sect ions," the groups of outside researchers who review grant applications.

Three things are notable about NIH clinical studies. First, it is a widely held opinion that most

[9] The PORTs take up only about one-fourth of the MEDTEP program. The program also includes investigator-initiated grants to examine specific issues relating to variation, outcomes, and methods development; and a program to support research centers on minority populations. Eleven such research centers are currently funded (821 ).

## BOX 4-6: The History and Components of the National Institutes of Health

The National Institutes of Health traces its origins to a single small laboratory, the Laboratory of Hygiene, that was established in 1887 within what was then the Marine Hospital Service (eventually to become the Public Health Service), for the purpose of investigating infectious diseases such as cholera The laboratory officially became the National Institute of Health by congressional fiat in 1930 and proceeded to undertake basic research into such widespread health problems of the day as tooth decay, undulant fever, and pellagra The National Cancer Institute was established separately by legislation in 1937 and for many years was functionally separate from the National Institute of Health

NIH formally became the National *Institutes* of Health in 1948, when four new institutes were created to work on heart problems, dental research, microbiological studies, and experimental biology and medicine Construction on the NIH clinical center, to further efforts to test the clinical applications of research, was begun at this time as well Additional institutes and centers were added over the following decades In 1994 NIH comprised the Office of the Director (which includes the Office of Medical Applications of Research) and *24* institutes, centers, and divisions

- National Cancer Institute
- National Eye Institute
- National Heart, Lung, and Blood Institute
- National Institute on Aging
- National Institute on Alcohol Abuse and Alcoholism
- National Institute of Allergy and Infectious Diseases
- National Institute of Arthritis and Musculoskeletal and Skin Diseases
- National Institute of Child Health and Human Development
- National Institute on Deafness and Other Communication Disorders
- National Institute of Diabetes and Digestive and Kidney Diseases
- National Institute of Dental Research
- National Institute on Drug Abuse
- National Institute of Environmental Health Sciences
- National Institute of General Medical Sciences
- National Institute of Mental Health
- National Institute of Neurological Disorders and Stroke
- National Institute for Nursing Research
- National Center for Human Genome Research
- National Center for Research Resources
- Clinical Center
- Fogarty International Center
- National Library of Medicine
- Division of Computer Research and Technology
- Division of Research Grants

SOURCES D Pugh, *The National Institutes of Health A Bethesda Landmark Celebrates Its Centennial* (Rockville, MD U S Department of Health and Human Services, Public Health Service, National Institutes of Health 1987) U S Department of Health and Human Services, Public Health Service, National Institutes of Health "Origins of the National Institutes of Health " pamphlet Bethesda, MD, undated U S Department of Health and Human Services Public Health Service, National Institutes of Health, Investment for Humanity, " Bethesda MD 1993

NIH RCTs focus on new technologies rather than existing therapies. Second, NIH does sponsor at least some comparative clinical studies of existing therapies. Third, however, there is at present no way to know the extent to which this is actually true. because NIH's data on its own RCTs are not complete enough nor detailed enough for anyone to examine the question.

As one example of clinical trials ongoing at one Institute. the National Eye Institute documents 21 ongoing studies (853). At least eight of these studies compare two or more technologies already widespread before the study began. NEI trials may be unusual in a number of ways (e. g., a substantial proportion address surgical interventions). Still. the} demonstrate that NIH does sponsor a presumably small but possibly significant number of comparative clinical trials of existing technologies. Even if such trials comprise only one-tenth of NIH clinical trials budget of not quite $900 million, as a federal financial commitment they would surpass the entire MEDTEP budget ($78 million) of AHCPR.

Other than NIH sponsorship of some comparative clinical trials that focus on existing technologies and broad populations. the NIH activities most directly tied to effectiveness research arc its development of health status and quality-of-life measures for certain diseases, and database resource activities. During the past six years, for example. the National Cancer Institute and the National Eye Institute (NEI) have held workshops on quality-of-life assessment in their respective areas, and the National Center for Nursing Research held a conference on methods to measure the effectiveness of nursing practice. A number of Institutes also maintain disease and procedure registries, a resource for researchers interested in augmenting databases (e.g., the Huntington Disease Research Roster, and the Vascular Surgery Registry. )

A notablc effectiveness research resource activity is a collaborative effort to link Medicare administrative data on patient scr}'ices with cancer epidcmiological data from NC I SEER (Surveillance, Epidemiology and End Results ) registry. a database that collects detailed, verified clinical data on persons with cancer in 11 areas across the country. The resultant merged database includes both data on tumor size and cancer severity and data on clinical services received by Medicare beneficiaries with cancer, as well as information on the costs of those seri'ices (610). The linked HCFA-SEER database will be used, for example, for a study of the patterns and outcomes of cancer care in the Medicare population (796). SEER data have also been used in AHCPR cancer outcomes studies (479).

### *The Centers for Disease Control and Prevention*

In 1946, the Office of Malaria Control in War Areas was replaced by the Communicable Disease Center, whose primary goal was to reduce the transmission of venereal diseases from homecoming soldiers (828). Now the centers for Disease Control and Prevention, CDC has grown to encompass 11 individual centers and offices whose common goal is disease and injury prevention (box 4-7).

In accordance with its mission, CDC stresses the epidemiology of disease and the identification of new disorders. Legionnaire disease, toxic shock syndrome, AIDS, and most recently a new deadly outbreak of a previously unknown virus in the American southwest have all been traced, described, and studied by CDC scientists. The agency's role includes some public health and prevention-related research (e.g.. into infectious diseases), but it is at least as much a service sponsor as a research agency; much of its role is in funding prevention programs.

The agency began to emphasize "prevention effectiveness" in the early 1990s (750,754). The focus for this effort was the establishment in 1992 of a Prevention Effectiveness Activity. with its own chief, within CDC Epidemiology Program Office. CDC staff specifically intended this activity to parallel the medical effectiveness initiative, with CDC assessing the effectiveness of population-bawl prevention efforts while others (primarily AHCPR) assessed the effectiveness of

**BOX 4-7: The Components of the Centers for Disease Control and Prevention**

The 11 operating units that collectively make up the Centers for Disease Control and Prevention demonstrate the agency's focus on population-based preventive and environmental health They are.

- National Center for Chronic Disease Prevention and Health Promotion
- National Center for Environmental Health
- National Center for Health Statistics
- National Center for Infectious Diseases
- National Center for Injury Prevention and Control
- National Center for Prevention Services
- National Institute for Occupational Safety and Health
- National Immunization Program
- Epidemiology Program Office
- International Health Program Off Ice
- Public Health Practice Program Off Ice

SOURCE U S Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention,organizational chart unpublished document, Atlanta, GA February 1994

"medical procedures" (749). In one example of the influence of this activity, a recent CDC statement on suicide prevention included a comment on the lack of information on the relative effectiveness of different strategies to prevent suicides (837).

CDC's focus on "population-based" interventions leads it to emphasize environmental and behavioral programs (e.g., lead abatement and public safety campaigns), although it also supports population-based programs involving clinical preventive services (e.g., programs to increase the rate of screening for particular diseases). However, the distinction between "population-" and "individual-based" clinical preventive services is not always clear-cut. One major current study taking place in clinical settings is a public-private collaborative effort, in which six health maintenance organizations (HMOs) are sharing existing HMO data on "prevention strategies for assessing the effectiveness of prevention activities in an HMO and community setting" (836a). Possible services to be examined include mammography utilization, antibiotic treatment of otitis media, diabetes management and prevention, and the treatment and prevention of domestic violence (751,753).

CDC is also a major compiler of health care registries and databases, although most of its registries have not so far played a major part in effectiveness research. Its compilation of mortality statistics and population-based health indicators through the National Center for Health Statistics, however, are fundamental to many studies.

### *The Office of Disease Prevention and Health Promotion*

ODPHP, a small office located within the Office of the Assistant Secretary for Health itself, was created by statute in 1977. Allocated a budget of somewhat less than $5 million for fiscal year 1994, its purpose is to establish national public health goals and strategies to achieve those goals, to act as a clearinghouse for information on disease prevention and health promotion, and to coordinate departmental activities in these areas (Public Law 98-55 1). To do this, ODPHP undertakes such activities as monitoring progress to-

wards the goals of its *Healthy People* **2000** report, operating the National Health Information Center, and coordinating health promotion and prevention activities among federal agencies and between the federal government and nongovernmental organizations.

ODPHP also occasional] y undertakes activities to fill perceived gaps in prevention activities undertaken by other federal agencies. This office, for example, helped develop the dietary guidelines (the "food pyramid") subsequently promoted through the U.S. Department of Agriculture. It also produces the Surgeon General Report on Nutrition and Health and is currently in the process of a structured literature review to support recommendations for dietary fat intake (325).

For the most part, ODPHP's activities draw on the results of effectiveness research, rather than sponsoring or conducting research itself. Two of these activities are discussed in more detail later in this report: sponsoring the U.S. Preventive Services Task Force and the Cost-Effectiveness Panel on Clinical Preventive Services. and convening an interdepartmental discussion group on cost-effectiveness of clinical preventive services.

### *The Health Care Financing Administration*

HCFA's mission is to administer the Medicare and Medicaid programs, the two massive programs that provide health insurance to elderly, disabled, and poor persons. As part of that responsibility, the agency includes within it an Office of Research and Demonstrations. This office sponsors such activities as pilot projects of novel approaches to delivering care to its constituent populations, evaluations of demonstration projects, and research into new methods of paying for services.

HCFA's largest contribution to effectiveness research is its enormous Medicare databases, which include detailed data on hospital care, outpatient care, health care institutions, and other factors. The potential of these databases to be rich sources of information on care patterns and outcomes was a major motivation for the federal government's "effectiveness initiative" (65 1 ). Their main disadvantages for descriptive purposes are that they do not include much of the information researchers want to discriminate among patients with different levels of health need. they often cover only a small slice of an individual health care experience (e.g.. inpatient care), and the clinical progression of disease can be inferred only indirectly, as a consequence of the procedures recorded in the data.

To address some of these issues, HCFA is currently involved in two separate efforts to provide greatly augmented databases. One of these is the linkage of the SEER-Medicare databases. The other, the Medicare Beneficiary Health Status Registry, will create a new database based on a survey of a large sample of Medicare enrollees (766). The survey. a mailed questionnaire, asks beneficiaries about their current health status. health risk factors, and socioeconomic characteristics. The survey is presently being pilot-tested.

HCFA also sponsors some descriptive studies to document outcomes associated with particular conditions or particular care practices in the Medicare and Medicaid populations. One major set of studies, for example. is examining the appropriateness and outcomes of care provided to Medicaid patients for conditions such as pediatric asthma, complicated delivery, and hysterectomy (638). Outcomes of care in Medicare patients who have end-stage renal disease, and outcomes of care in patients who have had hip surgery. also fall into this category. A few other studies deal with patterns of care and the examination of particular outcome measures. Examples are studies of post-hospital outcomes and studies analyzing the application of mortality and hospital readmission (796).

### *The Veterans Health Administration*

The Veterans Health Administration, located within the U.S. Department of Veterans Affairs (VA), provides for much of the health care of veterans of the U.S armed forces. In addition to its hospital system. VA has a long-standing set of supporting research programs that encompass prosthetics, medical care, and health services re-

search. Its budget for these activities in 1992 was $858 million. with a small but significant part of that budget derived from NIH through interagency transfers.

As with HCFA, the primary purpose of the health organizations within VA is to assure that the population for which it is responsible (i.e., veterans) are covered for their health care needs. Unlike HCFA, however, VA delivers these services directly. Consequently, it has developed in-house the clinical research and health services research capabilities that outside of VA are carried out by NIH and AHCPR.

Most research protocols are conducted by VA staff at one of its 172 medical centers around the country (many of which are affiliated with local medical schools). The three VA programs that directly sponsor research into the effectiveness of medical and mental health treatments, as well as the design and development of rehabilitative devices and systems of care are:

- **The Health Services Research and Development (R&D) Service,** the VA's in-house analog to AHCPR, performs most of the organization effectiveness research activities. Health Services R&D sets research priority areas and encourages research into these areas (523). "Outcomes and effectiveness research" was one of these priority areas in 1993 and included a number of projects analogous to the kinds of studies being supported in AHCPR's MEDTEP (e.g., the development of measures of health status and a feasibility assessment for collection of outcomes data on VA patients). In addition, building on the Medical Research Service's Cooperative Studies Program, Health Services R&D is funding multisite projects on such diverse topics as cardiac surgery outcomes, the clinical and cost impact of clozapine treatment on refractory schizophrenia, and a multisite randomized trial of team-managed hospital-based home care (875a).
- **The Medical Research Service,** VA's internal analog to NIH, sponsors its biomedical research and clinical trials projects. Many of its clinical trials are limited to VA patients (mainly elderly and disabled men), but some are potentially of broader applicability. Clinical trials particularly likely to fall into this latter category are large trials that are cosponsored by other DHHS agencies. The VA has considerable history and experience in multisite clinical studies.
- **The Rehabilitation Research and Development Service** has no real analog in the Public Health Service. It primarily conducts basic rehabilitative research, specialized product development, and tests of treatment or device efficacy. However, it also conducts a few descriptive and comparative effectiveness studies of interventions in the area of rehabilitation (e.g., alternative rehabilitation therapies for patients with multiple sclerosis).

## Coordinating Research Activities

Congress has designated AHCPR as its lead agency for research on improving the effectiveness of medical care through the evaluation of existing technologies and practices. The agency has had some successes at doing so but has encountered some substantial barriers as well.

Successful examples of intra-agency research coordination include the establishment of six "work groups" that enable research personnel from its various PORTS to meet periodically and discuss methodological issues. such as the use of health status survey instruments, common to all of the teams. As hoped, there has also been some natural coordination between a PORT and a guideline panel on the same topic; for instance, the researcher who was the consulting methodologist to the cataract guideline panel was the principle investigator of the PORT (724a). Guideline panels have several times been influenced by previous or con-

current work done by PORT teams (e.g., the incorporation of the prostate PORT's work on patient preferences into the recommendations of the prostate guideline panel.[10]

Interagency coordination of activities is more demanding and less successful. Underlying mechanisms that exist for coordinating among agencies include:

- **Representation on other agencies' advisory** bodies. AHCPR's advisory committee, for example, includes the administrators of seven other health-related agencies or departments as ex officio members.[11]
- Formal observer status for planning groups or task forces. NIH's OMAR, for example, regularly convenes a group of representatives from the different NIH Institutes to discuss issues for consensus conferences and other concerns. A roster of designated observers from other agencies. including AHCPR. are also routinely invited to these meetings, although the actual attendees from any particular agency may vary from meeting to meeting (78).
- Formal interagency coordinating groups. There appear to be no formal interagency coordinating groups on effectiveness research itself. There is a group of representatives from DHHS agencies that meets regularly to discuss issues in cost-effectiveness methods, convened by ODPHP to complement the work of its advisory task force (see chapter 7).
- **Conferences and workshops with invited participants and observers from other agencies.** Both AHCPR and other agencies frequently sponsor conferences and workshops on specific topics, to which staff and researchers associated with other agencies are often invited as either participants or observers. One classic example was a 1990 conference on primary care research, which was not only attended by but cosponsored by AHCPR. two NIH institutes, two other DHHS agencies. and a private nonprofit research foundation (798). Such efforts require that staff have foreknowledge of other agencies, and departments, interest, and that they act on it, which is not always the case. At a recent workshop on how to include cost-effectiveness considerations in clinical guidelines, for example, it did not occur to AHCPR staff to invite staff from ODPHP. who had been involved for some time in an effort to improve cost-effectiveness methodology. ODPHP staff knew to attend only because they found out about the workshop second-hand (869).
- **Interagency solicitation of cofunding for a study.** This has probably been one of the most successful mechanisms. It led, for example. to a collaborative study of the management of acute ear infections in children. a randomized trial being cofunded by the National Institute for Child Health and Human Development and AHCPR. According to AHCPR staff, this collaboration came about after NICHD sought cofunding for the study from other agencies (824). AHCPR agreed to help fund it after requiring some additions to the study protocol. Among several other examples of cooperative funding include the National Institute of Mental Health cosponsorship of the schizophrenia PORT; AHCPR's input into a prostate treatment study at VA; and AHCPR's funding of a quality-of-life component to be added on to an NHLBI-sponsored trial of treatments for hypertension and high cholesterol
- **Informal contact between staff.** The director of AHCPR'S Office of Medical Effectiveness

[10] AHCPR has several times deliberately assigned the same medical condition to both a PORT and a guidelines panel. This dual attention is in part the result of the priority AHCPR staff have placed for both activities on high-frequency procedures and conditions that have correspondingly high costs to the Medicare program.

[11] The seven agencies and departments are the Food and Drug Administration, the Health Care Financing Administration, the Substance Abuse and Mental Services Administration, the National Institutes of Health, and the Center\ for Disease Control. in the U. S. Department of Health and Human Services; the Department of Veterans Affair\; and the Department of Defense (52).

Research, for example, was previously affiliated with the VA. Informal contact can be a particularly important mechanism for coordination. although it is unreliable over time if more formal mechanisms do not exist because it is dependent on individual staff.

At present, instances of significant, productive cooperation among agencies and their activities seem more the exceptions than the rule, with cosponsorship of studies one of the more successful mechanisms. These both enhance AHCPR's resources to do broader effectiveness research and, presumably, help that research address questions of interest to other constituencies as well. The examples, however, are few.

Despite the problems with coordination among agencies in effectiveness research, formal mechanisms to increase coordination can present their own problems. Numerous people with whom OTA spoke during this study expressed doubts about the usefulness of formal interagency coordinating groups that meet periodically, because the activity tends to be viewed as relatively unimportant by participants, and the individuals participating tend to vary over time. Mechanisms to formally notify agencies about each other's activities also are viewed with skepticism because they tend to be considered a bureaucratic burden that would simply increase paperwork and discourage actual activity. Thus, relying on these mechanisms to increase cooperation may not be effective unless they are very limited and very targeted to specific purposes.

## CONCLUSIONS

**The crucial question for the next stage in effectiveness research is how to address the gaps that currently exist in this research.** Some of these gaps include:

1. Improving the efficient production of systematic reviews of existing studies, to make the best use of past efforts at clinical evaluation and to help identify important areas for research. Possible mechanisms for improvement include increasing funding for meta-anal - ysis (e.g., through specific grants, PORTS, or the U.S. Cochrane Center); requiring investiga[ors proposing new studies to demonstrate. through references to meta-analyses, that the research is not unnecessarily redundant: and maintaining a commitment to establishing a comprehensive database of controlled clinical trials.
2. Conducting more, and more efficient, clinical trials that yield valid comparative information on existing technologies, with results directly useful to patient and clinician decisionmaking. Possible mechanisms for achieving this objective are encouraging the use of patient-oriented outcome measures in more NIH-sponsored clinical trials; establishing and maintaining a comprehensive database of ongoing clinical trials sponsored by the federal government (and, where possible, private industry); and investing in a 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 more comparative evaluations of newly introduced technologies. Possible mechanisms include offering incentives to manufacturers to conduct comparative studies; encouraging or requiring payers, including government insurers, to link health insurance coverage for new technologies with evaluation of those technologies; and expanding the government role in sponsoring comparative evaluations of new technologies.

**At present, attempts to address any or all of these gaps face two major barriers. First, expanding the funding of comparative effectiveness research requires either new resources, which are extremely hard to come by, or a shift of existing resources, which faces the substantial opposition of those currently benefiting from those resources. And second, at present, there is no federal agency within the U.S. Department of Health and Human Services that considers the funding of comparative clinical trials of existing technologies to be one of its major responsibilities.** Thus. although the feder-

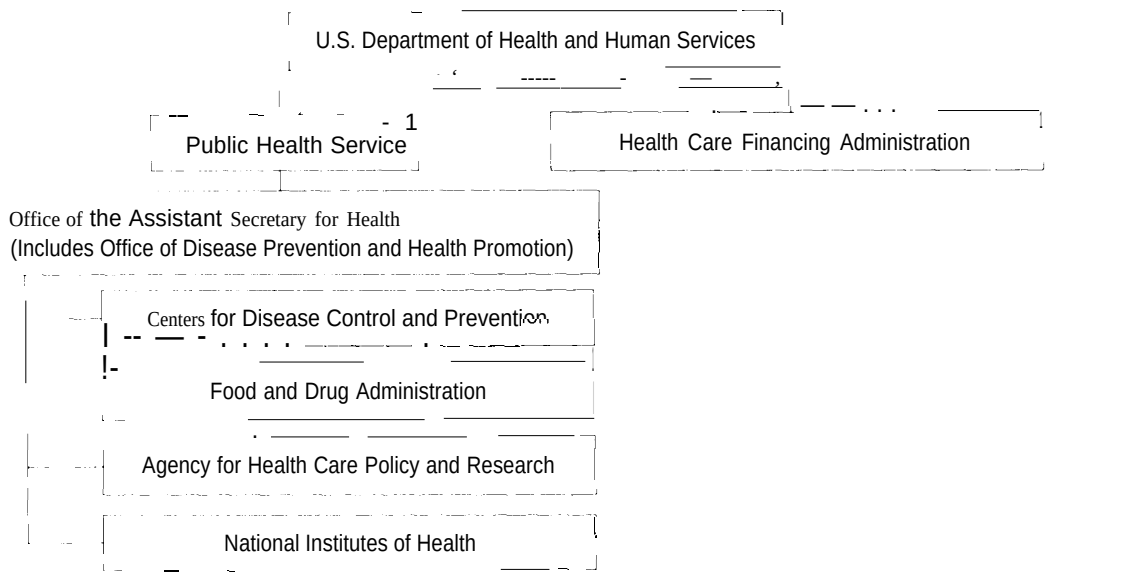
al government is investing resources in identifying high-priority questions about current medical practice, there is no real link between those high-priority questions and the actual comparative research that is being conducted.

The gaps in the federal government's current effectiveness research effort cannot successfully be addressed without assigning responsibility to fill them to a lead agency. Although in some respects the natural lead agency is AHCPR, its current level of funding is insufficient to fill these gaps unless it has the committed cooperation of larger agencies. At present, AHCPR has a moderatc level of interest in comparative clinic al trials. but it has neithcr the funding nor the leverage to ensure cooperation. NIH greater resources for conducting clinical trials are a natural target, but NIH does not view its role as primarily one of supporting evaluations of current technologies. Changing eithcr AH CPR's funding and leverage, or NIH's priori tics. will probably require congressional interest and intervention.

Although its trials arc not linked to AHCPR-generated research priorities, NIH does probably conduct a significant number of relevant clinical trials of existing technologies. The inability at present to identify relevant NIH trials, how its clinical trials funds are allocated, and other questions related to the characteristics of NIH studies deserves attention.

The VA is an under-recognized resource for federally sponsored effectiveness research. Although the population served by the VA has unique characteristics, many of the questions it faces are the same as those faced in the broader non-VA health care system. Consequently. the VA might well be a practical test in: ground for the potential to conduct effectiveness research and translate the results of that research into practice guidelines that can be implemented and evaluated. The VA is also well-organized for large, practice-integrated clinical trials. and for combining health services and clinical research aspects in single studies. Some examples of collaboration between the VA, AHCPR, and NIH exist, and greatcr collaboration might well prove worthwhile. Mechanisms for greater collaboration among these agencies, HCFA, and CDC regarding effectiveness research activities deserve emphasis and exploration.

**FIGURE 4-1: Relevant Agencies in the U.S. Department of Health and Human Services**




SOURCE Office of Technology 1994

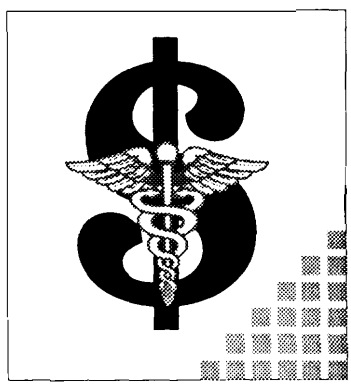
# The State of Cost-Effectiveness Analysis

5

Information about the costs and effectiveness of interventions is crucial to any decision about how to make resource allocation decisions in health care, an intrinsic concern of health policy makers. Purchases and consumers of health care services are increasingly concerned with not just the effectiveness but the value of the care being provided

Where both the costs and the effectiveness of health interventions are components of decisions, the quality and validity of the decisions can be increased by considering those components explicitly (780). Cost-effectiveness analysis can improve decision-making by forcing a structuring of the decision process and providing a framework for identifying and considering as many of the relevant costs and benefits as is feasible (780).

Cost-effectiveness analysis and related techniques are increasingly commonplace in health care discussions and literature. With the greater use and acceptance of this technique, however, old issues relating to its validity and the quality of studies that employ it have gained new importance. and new issues have arisen. This chapter reviews some of the changes in how and why cost-effectiveness analyses are done, and the importance of those changes to policy makers and other users of these analyses.



## ANALYTIC APPROACHES

The economic evaluation of health care alternatives comprises several related but distinct types of analyses. Three of these are

particularly relevant to resource allocation decisions and decisions about the relative value of alternative health interventions. [1]

**Cost-benefit analyses** are the oldest form of comparative economic evaluation. and they are frequently used in fields such as engineering and defense (box 5-1 ). In health care, cost-benefit analyses enumerate and compare the costs and benefits that arise as a consequence of applying an intervention (e.g., a medical technology or a public health program). Both costs and benefits are measured in dollars, enabling the analyst to compare a summary measure, such as the cost-benefit ratio or net cost (or savings), across any number of interventions (780).[2] Cost-benefit analysis is not a primary focus of this chapter, because its need to place dollar values on lives has resulted in disfavor among medical analysts, and it is relatively little used for the direct comparison of particular medical technologies. It is being applied in the analyses of some health programs, however, as discussed towards the end of this chapter.

**Cost-effectiveness analysis (CEA),** in contrast to cost-benefit analysis, does not convert units of health outcome into their "worth" in dollars. Rather, it calculates the cost per specified health effect of a technology or program--e. g., cost per lives saved, or cost per cases of' cancer avoided—and compares this cost-effectiveness ratio with ratios from other interventions (780). It is a crucial tool in the full assessment of medical technologies and services and is the primary focus of this chapter.

To be comparable in a CEA. all interventions must have their effects expressed as similar units, This is a problem when the goal is to compare technologies or services whose outcomes are not especially similar, such as prenatal care and stroke rehabilitation. Cost-utility analysis (CUA) is a variant of CEA that addresses this issue. In a CUA, the outcomes are expressed as uniform units of health that are presumed to have similar values across all conditions--"healthy days." "'health) years of life," or "quality-adjusted life years" (QALYs)—years of lifc saved by the technology, adjusted according to the quality of those lives (410).

Many analysts consider CUA and CEA to be separate entities.[3] In this chapter, however, cost-utility analysis is considered a subset of cost-effectiveness analysis that is especially powerful and that introduces some unique additional issues and concerns. The main reason for (his categorization here is pragmatic: both techniques are aimed at answering the basic question of the relative value of health interventions, and "cost-effectiveness" is the more inclusivc and familiar term. However. the North American research community also seems to be leaning towards considering CEA to be the umbrella term (867).

## USES OF COST-EFFECTIVENESS ANALYSIS

The purposes to which cost-benefit and cost-cfl'activeness analyses are applied by health care decisionmakers are increasingly diverse. Some potential uses are still more anticipatory rhetoric

---

[1] Other examples of methods of economic evaluation of health care programs include cost-of **illness** analyses, descriptive studies that are aimed at enumerating all of the costs related to applying a particular technology, or set of services. to the population" of interest (e. g.. people with cancer) (410) and cost-minimization analyses, which assess the least-cost method of achieving a particular outcome ( 1 83). Neither 01 these methods assigns relative values to both costs and outcomes across alter-rui( ive interventions, and neither i s addressed further here.

[2] Because both costs and benefits can be considered either positive or negative (e.g., a positive cost is also a negative benefit), the cost-benefit ratio is susceptible to manipulation (e.g., by restating a cost as a negative benefit and moving it to the denominator of the ratio). For this reason, the use of a summary measure of net cost (or net savings) is often preferred (217).

[3] In this framework CEA is used to compare the relative costs of achieving a single, common effect in alternative ways, with the effect usually calculated in natural units (e.g., lives). CUA is considered by those analysts to be a method of comparing costs of achieving either a single or multiple effects, where the value of the effects are specified in terms of their worth of a particular level of health status (187,394).

## BOX 5-1: The Origins of Cost-Effectiveness Analysis in Health Care

Attention to calculating the costs and benefits of public projects was a mandate of the U S Army Corps of Engineers, which was required by the River and Harbor Act of 1902 to assess the cost and benefits of river and harbor projects The Flood Control Act of 1936 took this concern a step further, requiring the U S Bureau of Reclamation to assess its water projects, stipulating that such projects should only be undertaken if the benefits should exceed the costs, As a result, cost-benefit analysis was applied to many water projects of the next *30* years (481) Other government entitles began using cost-benefit analysis to assist in program budgeting, and in the 1960s the Department of Defense began to employ the technique to evaluate alternative defense projects

Rice *(636)* was one of the first to employ methods for the economic evaluation of health care and health outcomes, in a 1966 paper on the cost of illness It rapidly became clear that this context raised considerable new issues and controversies Experts and stakeholder groups disagreed about what constituted appropriate outcome measures in such analyses, and valuing health and life in dollars—necessary for cost-benefit analysis—was controversial and, some maintained, unethical As a result the subsequent emphasis in health care tended to be on cost-effectiveness analysis, comparing health outcomes directly, rather than on cost-benefit analysts

During the late 1960s and early 1970s cost-benefit and cost-effectiveness analysts were applied to health subjects, but studies were not generally aimed at health declsionmakers who were making resource allocation decisions Neuhauser and Lewicki (560) helped change this situation when they analyzed 'the gain from the sixth stool guaiac " in a classic paper examining the marginal cost-effectiveness of a series of inexpensive tests for cancer screening.

In 1977 this initiative was followed by two milestone papers by Weinstein and Stason, which were published in the widely read /Vew *England Journal of Medicine* and accompanied by editorial discussion (720,906) These authors outlined the theoretical foundations and applications of cost-ef - fectiveness analysis of health care interventions and applied the technique to the allocation of resources for the medical management of hypertension Subsequent contributions to the health care cost-effectiveness literature in the 1970s emphasized the advantages of cost-effectiveness over cost-benefit analysis for analyzing health care topics (897) and highlighted the link between cost-effectiveness analysis and health technology assessment (244), These articles set the stage for a decade of growing interest and discussion of methodological issues and applications of the technique This literature, in turn, became the intellectual foundation for the increasingly widespread acceptance of cost-effectiveness analysis in every aspect of health care planning, management, and clinical decisionmaking that is coming to characterize the decade of the 1990s

SOURCE Adapted from B R Luce and K Simpson "Methods of Cost Effectiveness Analysis Areas of Consensus and Debate unpublished report prepared by Battelle Institute for the Pharmaceutical Manufacturers Association Washington DC Apr 22 1993

than reality, but they demonstrate the great variety of contexts in which CEA can be applied and the wide range of needs it can address. These uses include:

- manufacturers decisions about which lines of research on health care products to pursue,
- providers' decisions about purchasing new or replacement technologies.
- decisions by hospital or managed care formu - lary committees about which drugs to include on the list of pharmaceuticals that physicians may conveniently prescribe.

- payers' decisions about making changes in the way particular technologies are covered or paid for,
- inclusion in clinical practice guidelines, to help clinicians, patients, and payers judge the value of alternative care strategies,
- the design of insurance benefit packages under health reform, as policy makers attempt to judge which services should be covered,
- budgetary allocations within programs or agencies, and
- cross-agency resource allocation.

Although these uses differ considerably, in all cases decisionmakers can have a large stake in the results of the analyses they use. The validity and reliability of cost-ef'festiveness analyses, and differences in their usefulness that depend on the purpose to which they are put, are of widespread interest and concern.

## MECHANICS OF COST-EFFECTIVENESS ANALYSIS

A CEA typically proceeds through a series of steps, as follows (410,780):

1. **Identify interventions to be compared.** These may be different ways of providing the same technology (e.g., Pap smears), alternative medical treatments, or the variety of health programs across which society allocates its resources. The analyst specifies not only the exact technology involved but also the population it affects.
2. Identify costs of the interventions. These are the health care resources and other costs (e.g., travel and caregiver time) incurred when the interventions are used.
3. **Identify health and other impacts of the interventions.** This step requires identifying not only what health outcomes can occur in patients receiving the intervention, but often also identifying relevant impacts of the intervention on nonpatients as well (e.g.. on patients' employers and families).
4. Measure costs. After the analyst has identified what the costs are, he or she must apply "accepted economic procedures for attaching monetary values" to them (41 O). For example, if one resource cost identified is the need for a visit to a physician, the analyst must assign a dollar value to that visit+. g., the physician's charge for the visit, or the insurer payment for the visit.
5. Measure health and other outcomes. In CUA this measurement can be fairly complex, requiring the analyst to define a numerical health scale and using it to measure the outcomes. Thus, for example, if some of the outcomes identified as relevant are death, full health, and chronic disability, the analyst must assign values to these outcomes and the probabilities with which they occur.
6. **Examine uncertainties that underlie the analysis.** Rarely are many of the components of a cost-effectiveness model known with certainty. Usually they are either estimates, based on previously or concurrently collected data, with some confidence that the "real" number lies close to the estimate, or they are assumptions that are "best guesses" about the approximate size of the number. The effects of these uncertainties are usually examined through sensitivity analyses—"running the numbers" —again many times, with higher and lower estimates and alternative assumptions, to discover the range of possible results and the effects of different crucial assumptions on those results.
7. Present and interpret the results. The results of a cost-effectiveness analysis are usually presented as a single number whose magnitude can be compared across the alternative technologies analyzed—for example, the cost per life saved, or the cost per quality-adjusted life year. How those results are compared, however, can affect the conclusions a reader draws. A common form is to present the results as a marginal (or incremental) analysis, in which each result is compared with the previous one in stepwise fashion. This is the preferred method of presenting results inmost cases that involve analyzing alternative ways of reaching a common goal, because it allows the user to see what extra benefit is being gained for the extra costs of moving from one alternative to the next.

Questions about the relative cost-effectiveness of health interventions, like many other questions in health care studies. can be addressed either through retrospective or prospective studies. Most CEAs have traditionally been performed as retrospective analyses, using pre-existing data on the costs and effects of the different alternatives being compared, and a model created by the analyst that relates all of the components of the analysis. Models range from simple decision trees—schematic presentations of the choices the decisionmaker can follow and the consequences of each of those paths—to complex computer models that simulate the effects in great detail.

A cost-effectiveness analysis of Pap smear screening for elderly women (546) illustrates this common kind of CEA. In this analysis, the goal was to judge the effects and the costs to the Medicare program of covering this service, which was not previously a Medicare benefit. The population of interest comprised women age 65 and older. As a first step, the analysts chose to compare no screening with four other alternatives: one-time screening at age 65, and periodic screening beginning at age 65 and continuing thereafter every five years. every three years, or annually.

The main outcomes of interest in this analysis were the costs to the Medicare program and the health effects of screening for the women. These effects were measured in lives saved (i.e., deaths from cervical cancer averted). The model chosen for the analysis was a Markov model. a type of iterative decision model that represents the changes over time in the proportion of people who are in different health states (e.g., healthy, living with cancer, dead) based on assumptions about the probability with which people become ill, are cured using different treatments, or other factors.

In this analysis, running the model required estimates and assumptions of such factors as the total number of women in the population, the number who would choose to have Pap smear screens, the number of cancers detected by the screens, the cure rate for detected cancers, the death rate for cancers that are not cured. and the death rate for conditions other than cancer. It also required estimating the costs to Medicare of women undergoing screening, diagnosis, and treatment for cancer, At each iteration of the model. representing one year, the number of women alive and dead and the Medicare costs associated with prevention and treatment of cervical cancer were totaled. At the end of the iterations for that screening scenario (e.g., annual screening), the analyst summed the deaths per year and the total costs to Medicare.

Several of the model inputs required assumptions surrounded with a great deal of uncertainty. The sensitivity analysis. rerunning the model with slightly different assumptions. revealed that the overall costs and effects of screening varied considerably depending on the proportion of women assumed to develop cancer in the absence of screening, on the accuracy with which Pap smears detected cancers, and on the cure rates of treatments for detected cancers.

Finally. the analysts presented the results incrementally. Thus, it was possible to conclude that screening every five years was both less costly and more effective than one-time screening, under the baseline assumptions of the model. Screening every three years was slightly more expensive per life saved than screening every five years, but it saved many more lives. Screening every year saved the most lives of any scenario. but compared with the next-closest alternative (every three years) it was much more expensive for those additional lives that would be saved (546).

## ISSUES IN THE USE OF COST-EFFECTIVENESS ANALYSES

The literature addressing cost-effectiveness and cost-benefit in health care is no longer small. Over 3,000 articles and letters on the topic were published from 1979 through 1990 alone, of' which nearly 2.000 were analyses of particular interven - tions (219). (The remainder were methodologic papers, reviews, letters, and other forms of' commentary.) The growth of this literature has continued at a rapid rate since 1966 and show's no signs of slowring (219,781). Entire journals are now

dedicated to cost and cost-effectiveness analyses (e.g.,Pharmacoeconomics).

As CEA gains in acceptance and application to health care issues, both in the United States and abroad, the degree to which the results of these analyses are reliable and valid for the purposes to which they are applied has become a matter of increasing practical concern. Four issues in particular that affect the usefulness of this tool to policy makers are addressed here:

1. **The comparability of analyses:** the extent to which different analyses of the same question give similar answers;
2. **The quality of analyses:** whether the analyses incorporate the basic principles that most analysts agree would tend to increase the validity of the results;
3. **The increasing use of prospective cost-effectiveness analyses,** in which clinical trials are designed that measure costs as well as health effects; and
4. The implications of cost-utility analysis, and the basic assumptions that underlie the results of CEAs that express health effects in terms of patient "utilities" and quality of life.

## Comparability

The comparability of CEAs depends largely on the structure and assumptions of the analyses. As the number of analyses has proliferated, the approaches to analysis and the diversity of assumptions behind them have proliferated as well. As a result, CEAs in the current literature that appear to address similar questions often have results that in fact cannot easily be compared.

In a striking example, Brown and Fintor reviewed published CEAs of a single set of technologies: screening for the early detection of breast cancer (89). They identified 16 studies from 6 countries that assessed the cost per life- year saved or death averted. These 16 studies had wildly different results, mostly due to the different assumptions used in each cost-effectiveness model.

Four of the 16 studies reported breast cancer screening to actually save health care costs, while the other 12 found it to increase costs by differing amounts. Interestingly, those four studies all assumed that the lifetime costs of treating early-stage cancer were very different from treatment costs for later-stage disease, an assumption that Brown and Fintor traced to a single anecdotal report unconfirmed by any systematic study. The reviewers concluded that the results of those four studies were thus almost certainly invalid, and they turned to the other studies.

The cost-effectiveness ratios reported in these 12 studies differed enormously, ranging from $3,400 to $20,000 per life-year saved (89). When these results were adjusted to account for differences among studies in the assumption of the direct cost of screening (e.g., the cost of a mammogram), the range was even greater: $9,500 to $144,700 per life-year saved.

The reviewers then pointed out that the studies and their results were still not directly comparable for a variety of reasons. Two of the studies reporting the highest costs per life-year saved based their assumptions of the benefits of screening on evidence from short-term observational studies, and thus those analyses probably underestimated long-term benefits, which the observational studies were unable to detect. Two other studies were from countries with a relatively low prevalence of breast cancer, which would tend to make screening programs look less beneficial because they would detect fewer cases of cancer. And the studies differed in the exact technologies they were comparing; most compared mammography and/or clinical breast examination with no screening at all, while one study analyzed the cost-effectiveness of mammography compared to a pre-existing screening program using only clinical breast examination.

Setting aside all of the studies whose results were incomparable for these various reasons, the reviewers were still faced with two studies that ap-

peared to be very comparable in their methods and circumstances, yet whose results differed by a factor of four.[4] Examining these two studies more closely, Brown and Fintor identified a number of ways in which the underlying assumptions of these analyses differed. Examples include:

- frequency of screening (annually vs. every two years),
- efficacy of screening (30 vs. 33 percent reduction in mortality),
- length of time the population was "followed" and benefits were calculated (30 vs. 98 years), and
- cost of a mammogram ($40 vs. $50).

When the two models in these studies were rerun using the same numbers for these and other differing assumptions, the model yielded nearly identical results.

The CEAs examined in Brown and Fintnor's study exemplify many of the strengths and the weaknesses in the CEA literature. On the one hand, the similarity in the results of the final two analyses when they were run using similar assumptions affirms the basic reliability of the CEA method. On the other hand, the diversity of results from the overall set of studies would bewilder most users and might lead them to discount even the better quality analyses.

Mason and colleagues undertook a somewhat similar exercise and suggested seven categories across which cost-effectiveness studies often differed so much as to make their results incomparable. They analyzed a previously published table that compared the cost per QALY of 21 different medical technologies, ranging from advice to stop smoking to heart transplantation. The cost per QALY in this table ranged from 220 to 126,290 British pounds (499). Mason et al. examined the individual studies whose results were used to create this table and concluded that displaying their results in a single table was misleading. because the studies themselves involved such very different assumptions and circumstances that it was invalid to compare their results directly (495). Categories of factors that differed among studies were:

1. **Year the** study was done. In some older studies, the comparisons in the cost-effectiveness analysis were among technologies that were sometimes outmoded, making the results irrel - evant or misleading.
2. Discount rates. Analysts usually "discount" the value of costs incurred in the future, to reflect the fact that having a dollar today is generally considered preferable to having that same dollar sometime in the future (even if there is no inflation).[5] CEAs differ not only in the exact discount rate they use (e.g.. 4 or 6 percent) but in whethcr they discount future health benefits as well as future health care costs.
3. Preference values for health states. To calculate a QALY, the analyst "adjusts" years of life in a study population according to the level of poor health in that population and how bad people think those disabilities are—the "quality" of their lives compared with perfect health. Different ways of measuring these values give somewhat different results. and the QALY in one study thus may not actually be the same as a QALY in anothcr study. (This difference did not arise in the studies reviewed by Brown arid Fintor. because those studies used only likes saved rather than quality-adjusted lives as the "health effect" endpoint. )
4. Range of costs **included.** Most CEAs measure at least the direct costs associated with the technology being studied: the costs of applying the technology, the costs associated with side effects. and so on. Some. how'ever, also consider indirect costs, such as the monetary value of

[4] The two studies that appeared to be the most comparable were not necessarily of better quality than all of the excluded studies; they were just the most similar in method.

[5] Benefits and costs accruing now are worth more than if they accrue in the future. The procedure for adjusting for the time value of resources or costs is referred to as discounting, in that benefits and costs that are incurred in the future are valued in current dollars (906).

the patients' and families' time. Including indirect costs can change the results considerably.

5. Comparisons **used in the analysis.** Some studies analyze the cost-effectiveness of the technology of interest compared with doing nothing. Others compare it with another technology, while still others consider the incremental cost per QALY of expanding the technology to other groups of patients, or to other circumstances. In the breast cancer screening example above, this was clearly a problem in comparing results across studies.
6. Location and context of the study. Comparing the results of studies done in different countries without accounting for differences in the relative cost of alternatives within each country can lead to invalid results. In an example of how important country-specific prices can be, Drummond and colleagues compared the results of identically performed CEAs of a drug in four countries. They concluded that the drug was the most cost-effective in the United States despite being more expensive in this country than in others, because alternative treatments (e.g., surgery) were also more expensive (184).
7. **Quality of the study.** This category encompasses a wide range of differences among CEAs regarding both the assumptions they use and the way the results are presented. One criterion defining a poor-quality study, for example, is a study that does not identify the sources for its assumptions (e.g., about disease prevalence or health outcomes). Indeed, many published studies do not even state what their assumptions are. Where those assumptions are not stated, later users of the analyses cannot detect how those assumptions might have affected the cost-effectiveness results.

## Quality of Analyses

Differences in quality may affect the comparability of results, as described above. Of even greater concern, poor quality analyses may be invalid representations of the true relative costs and effects of the alternatives, possibly leading to worse rather than better decisionmaking, and very uneven quality of the analyses themselves.

Evidence suggests that the quality of CEAs is indeed cause for concern, even in the peer-reviewed literature. In an examination of this topic, Udvarhelyi and colleagues reviewed 77 articles on the cost-effectiveness of medical interventions that were published either in the late 1970s or the mid- 1980s and rated them according to six "fundamental principles" of cost-effectiveness and cost-benefit analysis (77 1). The principles they used were:

1. **An explicit statement of a perspective for the analysis should be provided. The** study should state whether, for example, it is the costs and benefits to patients, payers, or the health care system as a whole that is of concern.
2. **An explicit description of the benefits of the program or technology being studied should be provided.** This description should include assumptions about presumed benefits if the actual benefits are uncertain.
3. **Investigators should specify what types of costs were used or considered in the analysis.** The types of costs chosen should be linked to the perspective of the analysis (if the perspective is the hospital, for example, long-term morbidity costs might be excluded).
4. **If costs and benefits accrue during different periods, discounting should be used to adjust for the differential timing.** The reviewers did not require any particular discount rate, but they judged articles according to whether discounting was at least addressed whenever long time periods were involved.
5. **Sensitivity analyses should be done to test important assumptions.** As part of this principle, the reviewers stated that all assumptions of the cost-effectiveness model should be explicit, so that if the authors did not perform sensitivity analyses on important assumptions readers could be alerted to the possibility that the analysis' conclusions could change.
6. **A summary measure of cost effectiveness or cost benefit (e.g., the incremental cost-effec-**

**tiveness ratio of each alternative analyzed) should be calculated.** The reviewers specifically preferred the use of incremental cost-effectiveness ratios to average ratios.

Fewer than one-fourth of the analyses adhered to at least five of these six principles, and only half the articles adhered to more than three (771). The first principle was especially poorly followed: "Few studies (18 percent) explicitly stated what perspective the analyses used, leaving this to the readers' judgment. " Only 30 percent of authors used sensitivity analyses to test the vulnerability of their conclusions to key assumptions.

Equally troubling, the quality of analyses did not seem to be improving over time. The articles published in the later time period studied (1 985-87) were of no higher quality, overall, according to these criteria than articles published in 1978-80 (77 1). Yet the principles used in this review were not new ones (780), and previous reviews of the medical cost-effectiveness literature had also criticized the quality of papers published (18 1,894).

A recent review by Jefferson and Demichelli supports these general conclusions. In a very comprehensive review of the world literature on the cost-effectiveness of vaccines against hepatitis B, these authors found problems with both the comparability and the quality of studies (394). Of the 90 studies they examined, for example, only 37 contained a sensitivity y analysis, and only 36 clearly explained the time span over which benefits and costs were assumed to accrue. Of these, the time span used ranged from 4 to over 22 years. Similar-1 y, only one-fourth of the studies specified the discount rates they used, and the rates given ranged from 3 to 10 percent.

As the criteria used by Udvarhelyi and colleagues demonstrate, much of the concern about the quality of CEAs is not only that the appropriate structure and assumptions be used, but that these be explicitly stated. If the discount rate used in an analysis is not mentioned in the published account of the analysis, for example. it is both impossible to tell if an appropriate rate was used and impossible to compare the results of that analysis with others.

Explicitly stated structures and assumptions can enable later reviewers to redo analyses with common assumptions and examine the reliability of results, as Brown and Fintnor did in the case of breast cancer screening. But as CEAs become increasingly used in health care decisions. the inconsistency among analyses in their basic assumptions is itself of concern. End users of analyses may have neither the time nor. often, the training to examine and reconstruct individual analyses in detail.

Inconsistencies exist even among good quality analyses. One of the clearest and most long-standing examples of inconsistencies is in the use of discount rates, the analyst attempt to account for differences in how costs and benefits are valued in the future compared with the present. Virtually all analysts agree that costs that will not be incurred until far in the future should be discounted to reflect their true current value to decisionmakers in the present. Most also agree that health benefits received in the future should be discounted as well. What they do not agree on is exactly what that rate should be in either of these cases (399, 42 1,433,906).

Similar inconsistencies occur across other basic attributes. For example, there have been considerable strides in understanding and enabling the measurement of peoples' values for different health outcomes, but there is also still considerable disagreement over which measures are best. The diversity of the academic debate is reflected in the diversity of assumptions underlying different CUAs. QALYs are probably the most common health outcome measure, but measures such as the saved young life equivalent and healthy life expectancies have also been both proposed and used (567). And the analysts who calculate their results in terms of QALYs use a wide range of underlying methods to value those years, itself a source of considerable differences among studies (254,41 O, 568).

## Prospective Analyses: Clinical-Economic Trials[6]

As interest in cost-effectiveness has grown on the part of purchasers and payers of health care, the cost implications of new technologies are sometimes evaluated even while the efficacy of the technology is itself still being investigated in a clinical trial (box 5-2). Both the economic and clinical data from the trial are then analyzed to provide information about the cost-effectiveness of the technology. These *clinical-economic trials* are frequently initiated early in the development of a technology or medical practice (e.g., prior to FDA approval), although some have been conducted later, after the technology gains wider use in routine clinical practice.

CEAs based on economic data collected in clinical trials are still relatively rare (5), but they appear to be increasingly attractive. To date, most such studies have been sponsored by private manufacturers, whose goals for doing them include:

- to ensure that data on the economic implications of the technology are available for marketing purposes,
- to facilitate establishment of a price that will provide adequate return on investment while maintaining the economic viability of the technology,
- to formulate priorities on those drug indications, among alternatives, that should be sought in the FDA approval process, and
- to minimize the time between a technology's development and its coverage by health insurers.

Although manufacturers clearly have their own interests in supporting clinical-economic trials, prospective collection of data on costs as well as effectiveness has certain broader attractions. First, clinical trials conducted for the FDA approval process provide an opportunity to collect economic data when most needed to plan and guide appropriate utilization of technology by health care providers. Second, in contrast to clinical trials, studies using secondary data often incorporate data from disparate and at times incompatible sources, making the results difficult to interpret or apply. Thus, early, relevant, rigorously collected economic data coupled with a strong experimental design could be especially valuable.

Despite their considerable theoretical benefits, clinical-economic trials have significant drawbacks as well. Clinical trial timing, protocols, settings, and the nature of cost measurements all place their own special constraints on the collection of relevant cost data and have important implications for the validity and general inability of the results of CEAs based on these trials.

### *Influence of Clinical Protocol on Resource Use*

Clinical trial protocols themselves can influence resource utilization and costs. Protocols for clinical studies of a new technology often include tests to monitor study participants for serious or unknown side effects. For example, in the Women's Health Study conducted by NIH, in which hormonal therapy is being assessed, women will undergo frequent office followup. electrocardiogram, endometrial biopsy, and mammography to monitor safety of hormonal therapy. It is unlikely, however, that all of these tests will become components of the intervention in later routine clinical practice. Although such tests might be identified and their costs excluded in the data collection, monitoring can have more profound effects when an abnormal test induces further testing or treatment.

Another important component of many clinical trial protocols is "blinding" patients and physicians to treatment alternatives, to e1iminate biases in the perception of which treatment is preferred.

---

6 This section is based on N.R. Powe and R.I. Griffiths, "Economic Data and Analysis in Clinical Trials," paper prepared under contract to OTA, forthcoming, 1994.

## BOX 5-2: Examples of Clinical-Economic Trials

The most prominent use of clinical-economic trials has been in the area of new pharmaceuticals Most such studies are sponsored by the manufacturers, although they are often performed by academic institutions or other private organizations at the medical centers where the clinical study IS being conducted Examples include studies of the costs and health benefits of granulocyte macrophage colony stimulating factor (GM-CSF) as adjuvant therapy in relapsed Hodgkins disease (313), short- versus long-course antibiotic treatment of spontaneous bacterial peritonitis (665), and low- versus high-osmolall{y radiographic contrast media in patients undergoing cardiac angiography (611)

Cltnlcal-economic trials of medical equipment have also been conducted, although much less frequently Among the few devices that have been the subject of this kind of study are home air-fluidized therapy for treating pressure sores (733), automated versus manual methods of syringe filling (4) and videopelviscopy versus laparotomy for ectopic pregnancy (45)

A few cllnlcal-economic studies have been performed by providers to demonstrate the cost-effectiveness of one type of treatment over another in order to provide ]ustification for payment or to improve efficiency For example, a study conducted by the First Hill Orthopedic Clinic in Seattle demonstrated that in spite of the fact that patient length of stay was substantially more than that which formed the basis for Medicare payment, total hip arthroplasty in patients older than 80 was a cost-effective alternative to nurslng home placement (70)

Public agencies and private philanthropic organizations have also played a role in conducting or sponsoring clinical trials with an economic component For example, a study of outpatient management of burns using biobrane versus 1 percent silver sulfidiazine was sponsored in part by the World Health Organization (276) The National Center for Health Services Research (the predecessor of the Agency for Health Care Policy and Research) sponsored several clinical studies with an economic component, including an investigation of the costs and benefits of different long-term Immunosuppressive drugs in patients who had undergone kidney transplants (694). The National Cancer Institute has also recently showed some Interest in funding clinical trials with economic components (851 )

SOURCE N R Powe and R I Griffiths "Economic Data and Analysis in Clinical Trials " paper prepared under contract to Off Ice of Tee, hnology Assessment forthcoming 1 994

But the lack of knowledge about the treatment a patient has received can influence providers' resource use. For example, physicians may be more aggressive in managing a complication if they cannot be certain which treatment was uscd, since knowledge of the treatment influences their judgments about the importance and likelihood of complicat ions.

### *Cost Variability and Sample Size*

According to researchers who have conducted clinical-economic trials. a major problem in such trials is that the costs measured are generally much more variable than the effects (870 ). In addition. the distributions of costs are typically skewed, with either a few persons who use few services (or

who do not utilize services) or a few persons who utilize a large amount of resources and incur high costs (612). Because of the large variance in costs, a clinical trial usually needs more participants to obtain a statistically significant result for costs than are needed to make conclusions about effects with equal confidence. Some researchers have suggested that in fact costs often need not, and should not, be measured in clinical trials to the same level of statistical precision as health effects (1 86,572,573).

### *Cost Measurement*

Some costs are simply difficult to measure, or difficult to measure consistently, during a clinical trial. Some hospitals and practice settings do not have the sophisticated data systems available to identify institutional or provider costs in detail. This may limit the ability of researchers to perform studies in some settings or limit the type of data that can be collected (e.g., requiring the researcher to rely on charges rather than cost data). Multicenter studies raise important issues of standardization across different accounting systems and different types of providers.

Some costs may not be measurable at all during the trial because the economic consequences of treatment choices may extend far beyond the time horizon of that clinical trial. For example, thrombolytic therapy (drugs to break up clots that block blood vessels) used to treat an acute myocardial infarction can cause a stroke (a clinical endpoint), and the patient could require long-term nursing care, the cost of which could extend for many years. If the clinical protocol stipulated that patient followup end at the event of a stroke, the full economic consequences of the treatment choice would not be obtained through primary data collection.

### *Generalizability*

Economic data from a clinical trial may reflect "cost-efficacy" (2 18) rather than real-world cost-effectiveness, just as clinical trials may reflect clinical efficacy (whether the technology works under ideal or highly specified circumstances) rather than broader clinical effectiveness. Data from a study with strict selection criteria for patient enrollment performed at the best academic medical centers may be very different from data from a trial conducted in several community hospitals, which are more representative of the average U.S. hospital.

Often studies are performed at a single institution that is part of, or affiliated with, an academic medical center. Here, there are two possible problems that influence generalizability. First, medical practice may not be similar to that in nonteaching hospitals and other clinical settings. For example, physicians at teaching institutions may order more tests and consume more resources as a result of teaching or research activities. This may result in an overestimation of costs. Alternatively, if this more careful testing actually prevents complications of treatment, the data may lead the analysts to underestimate average long-run resource use.

Institutions that adopt technologies early may have the most experience in applying them. This experience could result in physician selection of more ideally suited patients to receive a treatment, or in better identification or management of side effects. This might not only translate into better outcomes than are realized in general medical practice (61 2) but also into more efficient use. In addition to differences in physician practice behavior, there can be differences in the cost of resources across institutions of difference sizes (economies of scale or scope), location (geographic variation in resource inputs) and organization characteristics (for-profit vs. not-for-profit institutions).

## Underlying Assumptions of Cost-Utility Analysis

The great attraction of cost-utility analysis is that it incorporates into a single outcome measure both the quantity and quality of life, without needing to assign any particular dollar value to that life. Thus, in theory, it can be used to compare the cost-effectiveness of technologies as diverse as treatment for ectopic pregnancy, whose primary goal is

to save lives. with hearing screening, whose primary goal is to increase the quality of life.

Unlike other CEAs, cost-utility analyses quantify not only the costs per relevant outcome, but the value to be placed on that outcome. In doing so, they raise issues that do not arise, or are not of as great concern, in other cost-effectiveness analyses. These issues revolve around the implicit assumptions that an analyst makes when he or she presents results as a cost per QALY or some other related measure. Because these assumptions are implicit rather than explicit, it is especially vital that policy makers, and others using these results, be aware of them.

### *Measuring Utilities*

A "utility" is a quantitative measure of the strength of preference for an outcome. The theory of utility measurement (the "expected utility theorem") states that in choosing among different options involving different outcomes (in this case, different health interventions involving different health outcomes), an individual should take the action that maximizes the expected utility of the outcome (41 O).

The most rudimentary method used by researchers to value outcomes is simply assigning arbitrary rates to a ranked list of outcomes. For example, death might be assigned a value of O, recovery with disability a value of 5, and full recovery a value of 10. A 1987 review of nearly 200 medical publications that included decision analyses found that a majority of analyses published up to that point valued outcomes either according to an arbitrary scale such as this or used even simpler measures, such as life expectancy or the simple occurrence of an event (i.e., the outcome measures of traditional CEA) (41 7).

For more complex measurements of utilities, one of the most commonly used is the rating scale, in which an individual rates each outcome by placing it at some point on a scale between two "anchors" (e.g., between O, representing death, and 100, representing perfect health). The Quality of Well-Being Scale, described briefly in chapter 4 as a tool for measuring health outcomes, uses such a rating scale to assign numerical values to health outcomes. Other methods of measuring utilities include techniques such as asking people how they would choose between one alternative that would lead to certain illness and another alternative in which one might either live in full health or die (the "standard gamble approach").

Academic debate continues regarding which valuation methods most accurately represent true utilities, as required by the theories that underlie decision analysis (762). Equally important to this debate, it appears that different measurement instruments administered in different ways can come up with substantially different results (254, 462,463,742). Methods that ask people how they would feel about having various disabilities or states of health for a short time (e.g., one year) may elicit different relative values for the different health states than a method that asks people to value those health states as if they were to have them for the rest of their lives. A method that asks about the (negative) value of losing an ability may yield different relative values than a method that asks about the value of gaining that same ability. There is some evidence that even a single measure can yield different utilities in different ethnic or geographic subpopulations (5 12).

The issue of different ways of measuring utilities has implications for both the comparability and the validity of CUAs. Two analyses of the same question may arrive at different answers using different measures of utilities. yet if both analyses express their results in QALYs, users maybe unable to pinpoint the reasons for the differences. The lack of agreement on the "best" measure of utilities means that measures vary considerably. Most importantly, from the user's perspective, there is no way to know how much this variability matters, There is no literature on the robustness of CUA results depending on the measures used, or on whether results are more sensitive to the measure used for some questions than for others.

Other. deeper issues of utility measurement are still open questions as well. For example, utility theory assumes that there is some simple underlying mathematical] relationship between how one

values short-term and longer lasting states of health (e.g., pain). But there are apparently no empirical studies to confirm whether this is true.

### *Whose Utilities To Measure*

In the simplest case of an individual making a personal decision, the utilities in a decision analysis are simply those of the decisionmaker. The individual places his or her own values on the outcomes and calculates the overall values of the alternative possible decision paths accordingly.

For other decisions in health care, however, the decisionmaker is generally making the decision on behalf of a group, and the preferences of that group must be considered. Much of the recent research into documenting health outcome preferences has been conducted in the context of specific clinical conditions (e.g., cancer or renal disease). In these cases, the preferences measured are those of patients with the relevant condition. One current Patient Outcomes Research Team (PORT) effectiveness research project, for example, is surveying stroke patients to document their preferences for the various health states and outcomes associated with stroke and conditions to prevent and alleviate it (496). Thus, a CUA that compares two management strategies for stroke could be reasonably sure that the values of the relevant group of patients were taken into account.

The least consensus regarding whose preferences to include appears for decisions that cross the boundaries of specific patient groups, particularly those that involve resource allocation decisions. In a notable and very public recent example, the State of Oregon proposed basing its Medicaid benefits package on, among other factors, the relative value of providing various categories of health services (788). The calculation of these values included measurements of the public preferences of various health states that might be affected by the services. The group whose preferences were surveyed were a sample of the general public who owned telephones. One group of critics argued that the appropriate group to survey would have been low-income persons, who were most affected by the program. A second group of critics argued that for valuing services to persons with chronic conditions, the preferences of persons who had those conditions—not the values of the general public—should be used.

The issue of whose preferences should be measured is important because several studies have shown that patients' preferences can differ substantially from those of healthy persons, health care providers, and other groups (63,75a,387). For policy decisions that involve broad resource allocation, there is a strong intellectual argument that general public values are the relevant ones. The experience with Oregon's waiver proposal, however, suggests that the public itself is not entirely comfortable with the implications of this approach.

### *Distributional Considerations*

The purpose of CUA in health care, like any CEA, is to improve decisionmaking that involves the allocation of health care resources. Unlike other CEAs, CUA actually incorporates values about health care outcomes into the quantitative part of the analysis. CUA's ability to assign numbers to the relative worth of very different health care activities is one of the attributes that makes this analytic tool attractive for policymaking. In doing so, however, it **introduces two dangers: decisionmakers may not fully understand the implications of the values that CUA incorporates; and they may not fully realize that although some values are quantified and incorporated into CUA results, other equally important values are not.**

The issue of whose utilities are represented in the CUA is one example of the importance to policymakers of understanding the assumptions behind it. Like any aggregate measure, the use of utilities in CUA implicitly assumes that it is average group values, and not individual values, that are important. Whereas the measured average utilities of a group (using a particular measurement instrument) tend to be quite consistent for any given question (763), measurements among individuals can vary enormously (667), and the value for

any one individual may be very different from the average value used to calculate QALYs.

This assumption probably makes sense for most policy decisions, because they involve the allocation of resources among groups. Nonetheless. it means that there are potentially large numbers of individuals who will disagree strongly with the decisions and their consequences, because the average group values do not represent their views. It is even conceivable that in some cases the number of individuals for whom average values are not representative will exceed those for whom they are.

Even if the disagreeing individuals are in the minority. their opinions may be of concern to public policy makers. For example, because CUA specifically assumes that quality as well as length of life matters, it places relatively less value on an extremely "low-quality" life (e.g., someone with only lower brainstem functioning) than on one of higher functioning. Individuals who strongly believe in the prolongation of life for life's sake, therefore. may sometimes find themselves at odds with the policy implications of some CUA results.

In addition to assumptions about whose values are important in distributing resources, the utility measurements underlying CUA tend to assume that relieving very severe distress that is very temporary is much less important than relieving more prolonged distress. CUA also assumes that "for a given degree of suffering, those whose illnesses happen to be cheaper to treat will be treated in preference to those whose treatments are more expensive" (623). Both of these assumptions are intended consequences of the principles of CUA as a method of allocating resources fairly. Nonetheless, there may be circumstances under which these assumptions are of concern. and they are certainly assumptions of which the users of CUA should be aware.

The most important point about the distributional assumptions of CUA, however, is what it does not assume. Because QALYs and other similar measures assume that all healthy lives are of the same value, they ignore social issues of distribution. The calculations in a CUA do not take into consideration, for example. whether benefits are being received by one subgroup (e.g., people in Nevada), while costs are borne by another subgroup (e.g., people in Connecticut). Nor does CUA account for the fact that society may *want* to value certain lives over others for certain purposes+. g., the sick over the healthy or the poor over the wealthy. Consequent] y. **these social decisions must be made outside the quantitative framework of the CUA.** There may be many times where, no matter how "valid" and high-quality the analytic results, policy makers will want to deliberate] y choose an intervention with a high cost-per-QALY over an intervention with a low cost-per-QALY, because the costlier intervention redresses social imbalances or achieves other social goals.

The importance of other social considerations is evident from a study. conducted in Norway, that tested whether health care priorities implied from conventional methods of deriving information on individual utilities actually conform to directly obtained priorities. The study is very small. based on a nonrandom sample of participants, and not easily generalizable, but it also raises some very unsettling questions about apply in: the assumptions of CUA in social decisionmaking.

In this study, when individuals were asked to prioritize between two hypothetical patients being admitted to a hospital, the great majority of respondents responded that priority should be assigned by order of admittance, regardless of differences in the likely health or disability of the patients after treatment (569). The author points out that this response differs from the implications of some American research (589), and acknowledges that cultural preferences may account for the difference. Nonetheless. the study challenges the concept that results from CUA are useful for deriving priorities for social policy. It suggests that more direct tests of the relationship between CUA implications and directly derived priorities is warranted before CUA results, even when combined with additional considerations of social issues, can be used to infer social preferences among alternatives with confidence.

## FEDERAL COST-EFFECTIVENESS ACTIVITIES

The federal government sponsors a number of activities relating to CEA and development of related techniques, but in light of the growing interest in the technique among private and public policy makers, the level of this activity is surprisingly small. Much of it, in fact, relates to effectiveness research—the development of underlying tools and data (e.g., decision analysis and health measurement instruments)—rather than the sponsorship of specific analyses.

### Agency for Health Care Policy and Research

The Agency for Health Care Policy and Research (AHCPR) inherited a long history of collecting and analyzing data on the cost of health care, a legacy of the health services research focus of its predecessor agency (the National Center for Health Services Research).[7] Most AHCPR activities related to CEA have been funding efforts to advance the underlying methods and tools, especially outcomes, quality of life, health status and patient utility measures, and methods of data acquisition (596). Sponsoring CEAs themselves has not been a particularly significant part of AHCPR's activities.

The PORT projects have been one of the most visible mechanisms for relevant research. All PORTS have undertaken some research related to measuring costs and health outcomes, and several are apparently conducting formal CEAs.

Another vehicle has been AHCPR's recent $14.5 million initiative to fund studies of the outcomes of existing pharmaceutical therapies (522).[8] Several of these projects are examining comparative effectiveness or costs, although only one study is a full CEA (on the cost-effectiveness of special pharmacist-based counseling for asthma patients, compared with usual care) (596).

AHCPR also funds a number of relevant, smaller independent studies. Again, a number of these are methodological in nature; examples include a multimedia-based method to assess patient preferences for different health states and treatments, and a "Longitudinal Comparisons of Measures for Health Outcomes" project. A few, however, specifically include analyses of the costs and effectiveness of particular interventions-e.g., a CEA of gallstone lithotripsy, and the cost-effectiveness of community-based care for elderly persons (817).

The 1992 act reauthorizing AHCPR (Public Law 102-410) made two significant changes that were intended to affect the level at which the agency sponsored CEAs or included CEAs done elsewhere in its own assessments. First, the act required that the individual technology assessments produced in the Office of Health Technology Assessment (OHTA) must include CEAs wherever valid data exist to support such analyses. Second, the act required that in producing clinical practice guidelines, AHCPR must consider the cost of alternative medical practices being addressed in the guideline. One consequence of these requirements is that to meet them, the agency must develop greater expertise in cost-effectiveness techniques. which might lead to greater interest in supporting research on the topic.

### National Institutes of Health

On the whole, National Institutes of Health (NIH) appears to view itself as a source of new technologies and information on the efficacy of those technologies, with little reason to be involved in studies of the costs or cost-effectiveness of those technologies. At an agencywide level, the main departure from this stance is a small publication

[7] At the time it was abolished, the full name of this agency was the National Center for Health Services Research and Health Care Technology Assessment. Here and elsewhere, it is still referred to by the shorter title for simplicity's sake.

[8] The $14.5 million represents the costs, for all years, of the projects funded in the first year of this initiative.

whose goal is to show that NIH research can sometimes lead to reduced costs to the health care system (842).

As in many other ways, however, there is substantial variation among Institutes in the perceived relevance of this topic. The National Cancer Institute in particular has been active in both conducting analyses and developing and critiquing analytic methods. Its efforts include analyses of the cost-effectiveness of prostate cancer screening, the development of a detailed cost-effectiveness model for cancer screening generally, and a series of activities relating to the cost-effectiveness of mammography screening for breast cancer (846). NCI also recently sponsored a conference on economic evaluation during clinical trials (85 1).

The National Institute of Mental Health is notable for the number of studies it sponsors in which reimbursement is an issue+. g., the impact of differing reimbursement mechanisms for intensive case management. NIMH also sponsors a number of studies that examine the relative costs and effectiveness of alternative interventions, ranging from studies of pharmaceutical interventions to studies of community treatment programs (570,846).

At least two other Institutes, the National Institute on Aging and the National Heart, Lung. and Blood Institute, also have intramural and extramural experts in economic analyses with whom they can consult. Few other institutes, however, have more than three or four ongoing studies in which measuring costs and determining relative cost-effectiveness is a major focus (846).[9] For many of those cost-related studies that are being conducted, a motivating force for measuring costs appears to be the desire to show that the NIH -sponsored research will ultimately result in reduced treatment costs for the condition of interest.

NIH is sponsoring a few trials in which economic data collection is performed during the trial itself, although the funding for the economic evaluation has come from elsewhere (e. g., foundations and AHCPR) (61 2).

## Centers for Disease Control and Prevention

Many of the pragmatic analyses and applications of CEA are done at the Centers for Disease Control and Prevention (CDC). in the context of increasing the “value for money" of the Nation’s public health programs and interventions. CDC has sponsored a few such studies off and on for at least a decade (e.g., 129,347). but the level of activity has expanded in the past two years with the agency's “prevention effectiveness” activity.

The CDC prevention effectiveness effort is explicitly designed to increase the incorporation of economic analyses into agency decisionmaking (749). In support of this effort, the agency is currently developing a basic, but fairly detailed, resource “how-to” manual on decision analysis, cost measurement, and CEA (831 ). The manual is being produced in-house and is expected to be published in 1994 (752).

CDC’s economic analysis activity is not large. but it is more extensive and more integrated with agency activities than similar activities at either AHCPR or NIH. It is also unusual in that a number of its economic evaluations are actually cost-benefit analyses, with benefits measured in dollars rather than the more common CEAs. Examples of ongoing and recent intramural analyses, which are spread throughout the various centers, include:

- a cost-benefit analysis of strategies to prevent nosocomial legionellosis (Legionnaire Disease) (836a),

[9] This observation i\ based on responses to a letter sent by OTA to the various NIH institutes during the course of this study ( see appendix A ).

- a cost-effectiveness analysis of folic acid food fortification and supplementation (836a),
- a review of economic evaluations of HIV 10 prevention and treatment programs (836a),
- the comparative costs and benefits of testing and counseling services for HIV-infected patients (353), and
- the cost-effectiveness of different strategies to prevent streptococcal infections in newborns (537).

In addition, CDC sponsors a number of extramural studies.

At least one explanation for CDC's relative interest in economic evaluation is that its programs face clearly defined limited resources, and they often have goals measurable through population-based health outcome measures (750). Cost-effectiveness and cost-benefit analyses thus can be directly useful to decisions that involve finding the best way to achieve a particular program's goals. Even so, its activities are not extensive.

## Health Care Financing Administration

The Health Care Financing Administration's (HCFA) primary purpose is to administer the Medicare and Medicaid programs, providing health insurance for elderly, poor, and disabled persons. Its research activities have tended to focus around payment issues. It is clearly a major potential consumer of the research and assessments of effectiveness and cost-effectiveness done by others, particularly AHCPR. Its reliance on AHCPR for these evaluations results in somewhat of a conflict regarding cost issues, however, since on its part AHCPR does not want to be viewed as in HCFA's "pocket."

Consistent with its mission as a major payer of health care services, therefore, many of HCFA's studies of effectiveness and care patterns include a strong emphasis on measuring costs. Even more notable, of the 10 HCFA condition- or technology-oriented studies in its 1992 research status report that most obviously include both a cost and an effectiveness component, seven are on the topic of preventive services. These seven projects represent nearly $20 million in research (796).

The largest of these seven projects, which accounts for half of the $20 million, actually comprises a multiyear experiment being conducted at five different sites (and a cross-cutting evaluation of the experiment). At each of these sites, study participants (all of whom are elderly persons) are randomized to experimental and control groups in which the experimental group receives a comprehensive set of preventive services. Both the costs of providing these services and their effects on participants' health status are measured (796).

HCFA's emphasis on the cost-effectiveness of preventive services is deliberate. The original Medicare statute provided only for coverage for acute care. Services such as preventive care, outpatient drugs, and long-term care were specifically excluded. Congress has been interested in extending Medicare coverage to more preventive services, and over time a few (e.g., mammography screening) have been added. Interest in adding to this short list is manifested in legislation that requires HCFA to study technologies such as influenza vaccine and comprehensive screening. Because of concern over the continued increase in Medicare expenditures, preventive services are held to a higher standard—a demonstration that they are not only effective but, by some measure, cost-effective—than are new diagnostic and therapeutic services that need pass only administrative scrutiny to obtain coverage.

HCFA has long contemplated establishing regulations that would permit it to use cost-effectiveness as a criteria more explicitly in decisions regarding coverage for any technologies and services. The agency first considered drafting regulations on the topic in 1979 (708), and actually did publish proposed regulations a decade later, in 1989 (54 FR 4302). Agency anxiety about the re-

10 HIV is the human immunodeficiency virus, which causes AIDS (acquired immunodeficiency syndrome).

ception of the regulations, however. has delayed their being made final. At this point agency staff state that new proposed regulations would have to be published for public comment before they can go forward, an action unlikely to happen until the likely shape of health reform is clearer (708).

### Office of Disease Prevention and Health Promotion

A significant effort to further cost-effectiveness techniques is being sponsored by the Office of Disease Prevention and Health Promotion (ODPHP), a small office under the Assistant Secretary for Health whose role is to coordinate and augment the Department of Health and Human Service's prevention activities. One of the major activities of this office was the establishment of a Preventive Services Task Force, which was charged with developing recommendations for clinical preventive services.

More recently, ODPHP has convened a Cost-Effectiveness Panel on Clinical Preventive Services to complement the work of the Task Force. Cost-effectiveness is an issue in determining both whether to recommend a screening test for a population and in the periodicity of screening. the frequency with which an individual should be screened for a particular disease. The Cost-Effectiveness Panel first met in 1993. It is expected to produce a report in 1995 that will attempt to advance the state of the literature on cost-effectiveness of clinical preventive services. address methodological issues and provide a framework for consistency in methodological approaches, and stimulate the application of CEAs.

Despite its 1 ink with the work of the Task Force, the recommendations of the Panel may not find their way immediately into the guidelines deliberated by the Task Force, due to hesitancy among Task Force members about the possible perception in their audience that recommendations would be too heavily influenced by cost (868). Nonetheless, the Task Force is including a short chapter on cost-effectiveness in its forthcoming update of its preventive services guidelines (697).

ODPHP has also convened an interagency discussion group to facilitate coordination and information flow across staff in the different agencies, such as NIH and CDC, that do at least some CEAs. [11] Persons from these agencies also attend the Cost-Effectiveness Panel meetings.

## COST-EFFECTIVENESS ANALYSIS IN THE PRIVATE SECTOR

### Recent Trends

A notable trend, associated with both the intellectual growth of effectiveness research and the growth in the private demand for health technology assessment' has been the increase in evidence on cost-effectiveness being produced by. or on behalf of, health product manufacturers, particularly pharmaceutical manufacturers. In 1992, for examplc, 27 out of 30 pharmaceutical firms participating in a survey said that they had begun to study "outcomes" (953). In a survey a year later, the companies reported that they were expanding ● 'outcomes research*' staff and doubling the number of economic studies (632). Pharmaceutical manufacturers are major clients of a number of technology assessment consulting firms.

Evidence of this increasing interest is apparent from the literature as well. Articles examining the economics of health care technologies have been appearing with increasing frequency since the mid- 1960s, but the number has mushroomed in the last several years, particularly for pharmaceuticals (215,475). In fact, 1992 saw the introduction of a new journal. *Pharmoeconomics*. devoted specifically to studies of the costs and cost-effectiveness of pharmaceuticals.

In addition to their use of academic consultants and private technology assessment consulting

[11] The full list of agencies and offices represented on the interagency discussion group comprises AHCPR, CDC, FDA, HCFA, Health Resources and Serv Ice\ Administration, NI H ( Including several indivi dual institutes), ODPHP, Office of the Assistant Secretary for Health, Office of the Assistant Secretary for Planning and Evaluation, and Substance Abuse and Mental Health Sin-\ ices Administration.

firms, some pharmaceutical companies have entire in-house divisions that conduct economic analyses of their products. Glaxo Inc., for example, established its pharmacoeconomic research group, with a staff of 19, in 1988. Eli Lilly similar-1 y has a staff of over 20 persons and has been conducting cost-effectiveness studies for about 4 years (755).

The major impetus for assembling these resources and studies appears to be their perceived usefulness as a tool for marketing products to cost-conscious consumers. In OTA conversations with manufacturers and others involved in the field, the prevalence of controlled formularies at hospitals and in managed-care organizations was one of the most frequently mentioned reasons for the proliferation of cost-effectiveness research of pharmaceuticals (174,509,473a). A recent detailed study of technology evaluation by health care providers found hospital formulary committees to be relatively sophisticated in the kinds of evaluations of pharmaceuticals they performed, sometimes assessing not only information on patient outcomes and drug costs but total related hospital costs as well (474). In studies conducted by the American Society of Hospital Pharmacists, 58.4 percent of hospitals reported a well-controlled formulary system in 1989, up from 53.9 percent in 1985 (789). Of HMOs, 22 to 55 percent in a recent survey reported having some sort of formulary (789). And market observers clearly expect the trend towards more control over the roster of drugs that doctors can prescribe to continue (56).

In addition, manufacturers are at least sometimes employing cost-effectiveness research to make internal decisions regarding the allocation of research and development dollars and to decide whether to continue to pursue the development of a particular drug or class of drugs (473a,755). In some cases, CEAs of a drug are done as early as phase II clinical trials, while the safety and efficacy of the drug itself is still in question.

The interest in conducting cost-effectiveness studies is evident in the medical device manufacturing industry as well, although to a lesser extent. Some companies regularly conduct CEAs, particularly of high cost, highly visible or controversial devices (174,461 ,672). But others believe that rapid technology obsolescence in a device product line often makes cost-effectiveness data irrelevant (473a,755). In addition, hospitals do not generally consider comparative cost-effectiveness when making technology purchase decisions, limiting their evaluations to simpler financial assessments (474). Thus, the perceived importance of CEA seems to depend heavily on the type and anticipated lifespan of the device.

## Issues

The increasing production and use of CEA in the private sector has several implications for federal regulators, planners, and health insurers. Most critical are the questions of whether those analyses will be of good quality; whether they will be comparable; and whether they will be either intentionally or unintentionally biased towards the result that the sponsor of the analysis (e.g., the product manufacturer) would find most favorable.

At the moment, there is no particular reason to believe that the quality of CEA performed by the private sector is necessarily any different from government or foundation research. Indeed. the same consulting firms and academics frequently contract with both private companies and public entities. Nonetheless, there is considerable concern that private sponsors might choose the enumeration of resources or costs in a CEA to increase the chances of obtaining a desired result (e.g., by including indirect costs only when doing so would increase the chances that the sponsor's product would appear to be more cost-effective). This possibility not only would bias results but might make studies across different sponsors incomparable. In addition, reviewers examining CEAs published in the medical literature could face substantial publication bias, since unfavorable privately sponsored studies might not be submitted for publication (particularly if they were performed in-house).

The concerns about framing studies to increase the probability of obtaining the desired result, and

the potential tendency to publish only favorable studies, are not entirely limited to the private sector. Some analysts point out that public organizations are not immune from these temptations, either.

One response to these issues has been increased attention, in the United States and elsewhere, towards making qua] it y standards for CEA more explicit and more standardized. The federal government is sponsoring one such effort in the area of prevention, the panel supported by ODPHP described above. A privately sponsored effort is also ongoing, based at the Leonard Davis Institute at the University of Pennsylvania (746a). A primary goal of that group is to devise ethical guidelines to minimize the potential for bias. The group held its first meeting in July 1993. Several international efforts to standardize CEA are described in box 5-3.

Because manufacturers who sponsor CEAs do so in great part to aid in marketing, this activity also affects the regulatory responsibilities of the Food and Drug Administration (FDA). The FDA oversees prescription drug advertising and labeling. A group at FDA meets weekly or semi-weekly to review cost-related claims in drug advertising and to examine what substantiation exists for the claims. When they see a questionable claim in advertising or detailing material, the y request supporting data from the manufacturer (539).

Whereas marketing claims regarding simple comparative cost are generally straightforward—FDA simpl y requires the source for the claim to be cited (e.g., a widely available list of drug average wholesale prices)--claims of cost-effectiveness arc considered claims about relative effectiveness and held to a more rigorous standard (539). FDA staff report that cost-effectiveness claims are on the increase (539). To address the need to evaluate these claims, the agency has been developing more expertise in the area and is developing a series of guidance documents to outline for companies what is necessary to support a cost-effectiveness claim (539).

## CONCLUSIONS

Analyses that consider the costs and health effects of an intervention in a structured fashion can improve decisionmaking, and greater use of CEA in this context appears to be on the horizon. Although formal CEAs are still not in abundance. the number of analyses on health care topics has grown considerable y over the past decade. Similarly, CEAs are still not routinely applied to most health care decisions, but the sponsorship, the use, and the interest in these analyses has been increasing rapidly.

As the use of CEA increases, attention to the validity and comparability of analyses becomes doubly important. Despite the greater number of analyses being produced now than in the past, it is not clear that the overall quality of analyses has improved. Furthermore, **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 deserve attention and support.**

The greatest methodological change in the field in recent years has been the increasing prominence and sophistication of cost-utility analysis. The great attractions of CUA are that it incorporates a broader range of relevant health outcomes than simply lives saved, cases of disease averted, or other single measures; and that it quantifies these outcomes so that interventions with greatly differing purposes can be compared directly. Consequently. it has immense potential appeal to policy makers who are making allocation decisions across broad areas of health care.

**The greatest danger of CUA is that it incorporates so much into a single measure, such as the QALY, that policy makers may be unaware or less attentive to the many issues that CUA does not address.** CUA, for example, does 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

## BOX 5-3: Standardizing Cost-Effectiveness Analyses: International Activities

The uneven quality and Inconsistencies across cost-effectiveness analyses (CEAs) have led to a call for greater attention to standardizing cost-effectiveness methods (1 85) In fact, an international movement is underway to begin to do so in at least four countries and communities

**AUSTRALIA.** Under Australia's health insurance system, an independent committee makes recommendations to the government regarding coverage for new drugs In 1987, the committee was directed to begin considering costs as well as effectiveness in its decisions In 1990, the committee issued draft guidelines to encourage (and, ultimately, mandate) manufacturers to present data on the cost-effectiveness of their products when applying to have their products covered. To Increase the comparability of these cost-effectiveness data, the guidelines set some explicit standards that analyses must meet to be acceptable These include

- standard values to be used for certain health care units (e.g , physicians visits),
- discouraging the use of redirect costs in the analysis (e g., lost work productivity),
- permitting the use of some direct nonmedical costs (e. g , costs of home help) as well as medical costs,
- encouraging the use of outcome measures such as Improvements in functioning, and
- encouraging the use of cost-utility analysis where possible

Manufacturers are to submit the results as incremental cost-effectiveness comparisons of the new drug with either the alternate drug most widely used in Australia, or to the alternate drug the new pharmaceutical would replace (39, 182),

CANADA. The provincial government of Ontario and the Canadian national government are in the process of developing guidelines similar to those in Australia. The Canadian guidelines, however, are not regulations Rather, they are an effort to develop "consistent and uniformly understood principles, definitions, and methodologies for the conduct and evaluation of economic analyses. " These suggested standards are intended to assist drug manufacturers in meeting the growing demand on the part of provincial drug programs for cost-effectiveness information (1 60) The authors stress that the guidelines are to be flexibly applied. Some of the recommendations of the draft guidelines are.

- all "relevant" costs should be included (indirect as well as direct costs where appropriate),
- the analysis should state clearly its perspective (e.g., societal vs. insurer), and two separate analyses from different perspectives may be appropriate,
- outcomes are to be expressed first in natural units (e g. myocardial Infarctions avoided) and also in alternate units such as benefits (dollars) or utility (using, for instance, quality-adjusted life years (QALYs)),

measured by cost per QALY may in fact not always be the best allocation of social resources when concerns such as distributive justice are taken into account. Nor does it address the question of whose values should matter the most for particular decisions: it treats all values as the social average. The issue of whether a particular demographic minority's values should matter more than average public values, or the subpopulation from whom the values in the analysis were derived, in a particular case is a decision that is ultimately a political one.

**BOX 5-3 continued: Standardizing Cost-Effectiveness Analyses: International Activities**

- an explanation of equity assumptions should be included (e g , whether QALYs gained by all Individuals were considered equal), and
- the Incremental cost-effectiveness of one drug relative to another is to be expressed as a ratio of the cost to the outcomes (180,578a)

**SPAIN.** In recognition of a growing desire to use CEAs in resource allocation decisions in Spain, a group of researchers, working with the government, have proposed a set of standardized approaches on seven topics The topics covered by their working document are (659)

- selection of alternatives being compared,
- the Inclusion of direct vs Indirect and intangible costs
- the analysis,
- methods for valuing costs,
- measures of health effects used,
- time horizon and discounting,
- treatmeny of risk and uncertainty,
- range of sensltivity analyses, and
- presentation of the results

The goal of this group is maximum standardization, particularly in areas such as the range of assumptions tested in sensitivity analyses and the discount rates used. The effort was not complete as of 1993 but agreement on standardization for many of the topics was well underway (657)

**THE EUROPEAN UNION (EU).** Another group of researchers, including the head of the Spanish group, have proposed an undertaking similar to the Spanish one but on a larger scale This effort is only just underway, and funding (by the EU) was still tentative as of 1993 The goal of the joint European effort is an ambitious one: "to develop and propose a unified methodology for the economic evaluation of health technologies to be adopted by EU regulatory agencies, national administrations and European multinational companies operating in the health care field with the eventual aim of harmonizing regulatory practices across EU countries (658) The proponents of the proposal argue that standardizing methods across countries will benefit those carrying out studies and will improve the transferability of results of studies conducted in one country to other countries where decisionmakers face similar issues

SOURCE Off Ice of Technology Assessment 1994, based on sources as shown Full citations are at the end of the report

CUA also incorporates a number of implicit assumptions of which its users should be aware. Most fundamentally, it assumes that the values elicited from people about health in surveys translate into valid representations of their preferences for different interventions or resource allocations. The one direct test of this assumption suggests that it may be flawed (569).

The other very significant change in cost-effec - tivencss methodology is the growing practice of conducting CEAs simultaneously with early clin ical trials of a new treatment efficacy and safety. This practice raises questions about the sample size needed for the economic component of these clinical-economic trials, and whether such studies may be biased towards finding no difference in costs between treatments. 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 these potential concerns, cost-effectiveness studies and related activity in the private sector has boomed. Spurred by the need to deal with an increasingly sophisticated cadre of managed-care administrators who are very cost-conscious, private industry 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 CEAs 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 the FDA will need to become increasingly sophisticated at evaluating such claims.**

Given the growing level of interest among private and public policy makers alike in CEA, the **federal government's level of activity related to CEA is surprisingly weak.** Only in the area of preventive services is there any substantial investment. Although there is some overlap in activities that warrants close communication and coordination among agencies, ODPHP is well positioned to play this role and to support it with methodological work, as it is doing. AHCPR also supports some relevant methodological research, but in general CEA related to treatment and long-term management has been given relatively little attention by federal agencies.

Although there is still no uniform agreement among policy makers or the public about what role information regarding the cost-effectiveness of treatments should play in insurance coverage decisions, it is possible that there may be more agreement on this point in the near future. At present, federal agencies are not well positioned to support this research, through either their in-house expertise or their current sponsorship of methodological studies.

Perhaps the most ambitious endeavor regarding the use of cost-effectiveness information in health policymaking to date has been Oregon's attempt to use this information as a foundation for creating an entire health benefits program for the state's Medicaid program. The state's attempt to rank all primary and acute care services according to their importance, costs, and effectiveness gained national attention and spawned a furious debate over the ethics of the process (735). The ultimate reliance of Oregon on the opinions and judgments of its appointed commissioners to value services, and the lack of solid data to assist them in making their decisions (788), was a blunt reminder that the use of CEA in health policy decisionmaking has limitations.

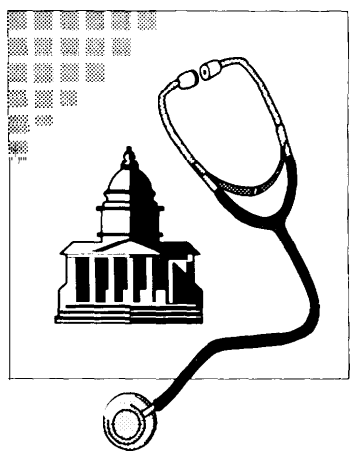
# The Federal Role in Health Technology Assessment

6

Health technology assessment has been a direct concern of the federal government since at least 1976, the year the Office of Technology Assessment (OTA) published its first report on the topic (778). Interest during the ensuing years has waxed and waned as Congress and other interested parties debated the appropriate uses of health technology assessment and the government's role in this activity,

The debate has been complicated somewhat by the diversity of activities that are sometimes labeled "technology assessment." Although in the context of health care this phrase has been defined comprehensive y at times, to include an analysis of the "evidence of [a technology 's] safety. efficacy. cost. cost-effectiveness, and ethical and legal implications" (597 ), it is also often applied to evaluations of only some of these components. Health technology assessment has been used to describe activities as diverse as hospital purchasing decisions (477a). randomized clinical trials (165), and the cost-effectiveness evaluation of public health programs (348). Indeed. the association of technology assessment with cost-effectiveness analysis (CEA) has led some researchers to consider health technology assessment and CEA to be nearly synonymous (270).

OTA's definition of technology assessment is broader and more policy-oriented than many of the uses of this term elsewhere. In this report. as in previous OTA reports, "health care technology" comprises drugs. devices, procedures, and the organizational and supportive systems within which health care is delivered (780). The inclusion of "organizational and supportive systems" is an acknowledgment that the implications of a health technology depend on its context, and that clusters of individual



technologies organized in a specific way can themselves become a more complex technology--e.g., an intensive care unit.

**"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 performed for the purpose of providing input to a policy decision.** Requisite components of a health technology assessment include the collection or generation of information about the technology (including, e.g., information about its effectiveness and cost-effectiveness); a synthesis and critical analysis of that information in the context of the policy decision being addressed; and presentation of the result in language that is relevant to the decision.

In this framework. the perspective and breadth of a given technology assessment is determined by the policy decision to be made. If the decision relates to insurance coverage, for example, the assessment might address issues of effectiveness, utilization, costs to the insurer, effects on the costs and use of other services, and potential for legal liability in the case of noncoverage. In contrast, an assessment of the same technology as part of a national research and development policy might place much more emphasis on factors that influenced the technology's development, and on the broad social consequences of its application.

CEA (discussed in the previous chapter) is often an important component of technology assessment, but the two activities are not synonymous. A CEA alone is only an adequate technology assessment when costs and effectiveness are the sole issues relevant to a policy decision, as might be true for a few clinical management or coverage policies. But CEA is a powerful tool for technology assessment, and interest in assessing medical technologies was a major impetus behind the initial development of the field (780). The two activities are clearly closely linked.

This chapter describes the federal government's involvement in health technology assessment and the relationship of technology assessment to clinical practice guidelines. It also describes the escalating private interest in health technology assessment.

## GROWTH OF HEALTH TECHNOLOGY ASSESSMENT

### Debating the Federal Role

As originally conceived, technology assessments were to be aids to public policy makers. The term itself was coined by legislators concerned about the social impacts of technologies (box 6-1). Despite this history, the federal government role in health technology assessment has been a subject of intense controversy from the beginning.

The earliest reports about health technology assessment (778,779) drew attention to the fact that most medical technologies were introduced and widely adopted without undergoing any rigorous evaluation. Few were adequately evaluated even for their safety and efficacy, much less their broader effectiveness, costs, and social implications. Although the Food and Drug Administration (FDA) evaluated evidence that drugs were safe and efficacious as part of its regulatory responsibi - lities, similar responsibilities relating to medical devices were enacted only in 1976 and were much more limited. No systematic process of evaluation of medical or surgical procedures existed at all. Nor has the FDA generally viewed its authority or responsibilities as extending to the examination of economic or social issues.

Because they largely escaped FDA scrutiny, medical devices and procedures were a natural first target for federal technology assessment efforts. At a time when rapidly rising health care expenditures were becoming a matter of increasing concern, the introduction and diffusion of expensive medical devices was considered a major contributor to medical costs (597). The federal government's support for health planning, and its financial interests in the Medicare and Medicaid programs, made devices such as the computed tomography scanner particularly attractive targets for assessment (see, e.g., reference 782). The association of technology assessment with the valuation of expensive devices for the purposes of government health planning and technology man-

## BOX 6-1: The Origins of Technology Assessment

The concept of "technology assessment" is rooted in the political and social debates o? the 1960s and 1970s, when the environmental and social consequences of technologies such as the pesticide DDT and the supersonic transport plane were prime topics for political discussion at every level Credit for Introducing the phrase is traditionally assigned to Emilio Daddario, former chairman of the Science, Research and Development Subcommittee of the House Science and Astronautics Committee of the U S Congress, who defined it in 1967 as" ...a form of policy research which provides a balanced appraisal to the policy maker It is a method of analysis that systematically appraises the nature, significance, status, and merit of a technological progress" (1 47)

Early uses of the term specifically required that technology assessments should identify indirect effects of technological innovations and assess these effects for the purpose of improving decisions regarding the social use of technology (428a,774). The idea that "technology in this context should be broadly defined was explicit, an early report to Identify candidate technologies for assessment included such items as acupuncture for pain relief, early tests for fetal deformities, and compulsory heroin treatment clinics (554).

The concern regarding the impact of technology on society led directly to the creation of a small legislature support agency, the Off Ice of Technology Assessment (OTA), to assist Congress in making decisions that revolved science- and technology-related issues The OTA Health Program issued its first report, on Development of *Medical Technology Opportunities for Assessment?,* in 1976 OTA continues to produce assessments of both Individual technologies and broader technology-related issues at the request of Congress

SOURCE Off Ice of Technology Assessment 1994, based on sources as shown Full citations are at the end of the report

agement, however, meant that from the beginning the activity was often considered directly counter to the interests of the health products industry and the autonomy of professional medicine.

The federal government took several steps in the 1970s to fill its perceived need for information about potentially problematic health care technologies. In 1972, Congress created OTA to perform technology assessments and related analyses for the purpose of assisting with legislative decisionmaking (Public Law 92-484). The Health Program was created within OTA in 1975 and released its first report, on opportunities for assessing medical technologies. in 1976 (778). OTA continues to perform assessments of health care technologies and technology-related issues, but because it is located in the legislative branch of the government, its role in producing health technology assessments is limited to those requested by Congress.

In 1977, NIH (at the urging of Congress) established the Office of Medical Applications of Research and its Consensus Development Program. Its stated goal was to bring together physicians, consumers, scientists, and others "in an effort to reach general agreement on whether a given medical technology is safe and effective" (864). The first Consensus Development Conference, on breast cancer screening. took place in September 1977 (866).

Then, in the following year, Congress established the National Center for Health Care Technologies (NCHCT) (Public Law 95-623). NCHCT had an ambitious mandate that embraced a broad role for the federal government in con-

ducting and facilitating health technology assessments. Its mandated activities included setting priorities for technology assessments, conducting assessments, developing standards for the use of technologies, and advising the Medicare program regarding coverage for new technologies (68, 241 ,598). Its mechanisms for carrying out this mandate included not only intramural staff analyses but clinical studies and the use of panels of expert advisors.

NCHCT was confronted with immediate opposition from medical and industry organizations (63 1). The establishment of standards particularly concerned these groups. During hearings preceding the reauthorization of the center in 1981, the American Medical Association (AMA) testified that NCHCT would interfere with the practice of medicine (76), and the Health Industry Manufacturer's Association argued that NCHCT's functions were unnecessary, would stifle innovation, and duplicated those of NIH (674,675). Congress did reauthorize the center, with an abbreviated role that eliminated the standards mandate. The administration did not request funding for NCHCT, however, and Congress elected not to appropriate the authorized budget.

After NCHCT's political demise in October 1981, a vestige of the center became the Office of Health Technology Assessment (OHTA) in the National Center for Health Services Research (NCHSR). The duties of this small office were reduced to advising the Medicare program regarding the safety and effectiveness of medical technologies being considered for coverage. OHTA assessments relied largely on staff-conducted literature reviews, surveys of other agencies' activities and evaluations, and on unpublished clinical evidence provided by manufacturers and others.

Although OHTA was responsive to Medicare concerns, it did not address other interests that were also initially behind the creation of NCHCT. States, private insurers, and federal policy makers with broader concerns in the social, economic, and health care implications of medical technology still lacked access to assessments that incorporated these concerns.

Congress created the Prospective Payment Assessment Commission in 1983 to address some of the federal needs for technology assessment in light of changes to the way Medicare paid hospitals (Public Law 98-21 ).} Unlike NCHCT, the Commission was supported politically by industry, which saw it as a way to have a voice in Medicare payment policies that would affect the adoption of expensive new technologies (363a). Over time, however, the technology assessment component of the Commission's work has declined, and this activity is now manifested primarily through efforts to assess the extent to which hospital payments should be changed to account for technological innovation (3,785).

Congress found a temporary home for broader efforts in 1984 with the establishment of the Council on Health Care Technology Assessment (Public Law 98-55 1), which was placed under the auspices of the Institute of Medicine (IOM). The IOM had previously been quite active in addressing the issue of medical technology assessment, including the development of a report on technology assessment and its component techniques (366). The Council was charged with operating a "clearinghouse" for technology assessments, conducting assessments, and furthering methodological development (259). Its contributions included a directory of organizations that performed medical technology assessment (367) and several publications on conceptual and methodological issues (368,374,375). The Council found it difficult to raise the private funds necessary to help support its activities, however (635), and its authorization was allowed to expire in 1989, the year that the

[1] Beginning in 1983, Medicare ceased reimbursing hospitals for their Medicare-related inpatient expenses on the basis of actual cost and began paying for them under a prospective payment system based on diagnosis-related groups (Public Law 98-21 ).

Agency for Health Care Policy and Research (AHCPR) was created.

In AHCPR's mandate, Congress re-established a more direct role for the federal government in conducting health technology assessments for brooder purposes. Along with its mandate to support effectiveness research, AHCPR inherited the old NCHSR[2] and most of its functions, including those of OHTA. To date, OHTA technology assessments have been done only in response to requests from federal health program policy makers—specifically, the Medicare program and the Department of Defense's CHAMPUS insurance program for military dependents (351). The output of the office is accordingly small, averaging fewer than five assessments or reviews (more limited evaluations) per year (box 6-2). The 1992 legislation reauthorizing AHCPR now permits the agency to perform individual technology assessments for more general reasons (Public Law 102-41 O), but whether the agency will have the resources and the desire to do so is still unclear.

## Health Technology Assessment in the Private Sector

For all of the federal government's 15 years of involvement in health technology assessment, it has never really carried out the central technology assessment repository function originally envisioned for NCHCT. A recent, briefly contemplated proposal to augment OHTA's funds with private funds and cater to a larger clientele—particularly the needs of private health insurers—was dismissed as politically and administratively infeasible (440,521).

But in the private as well as the public sector, the demand for timely and relevant technology assessments has increased. Rather than information on broad social implications or regional health planning efforts, private users of health technology assessment want targeted information to help them make coverage, purchasing, and management decisions. Ironically, the interest in assessing technologies in order to monitor and control their use remains a major impetus for the demand for this activity, but the planners are now often private managed care organizations and multihospital systems rather than governments.

Stimulated by this demand, a small but explosive private market for health technology assessments produced by and for health care providers. payers. and manufacturers is flourishing. In this market, activities have largely abandoned technology assessment initial emphasis on broad social and ethical impacts and focused instead on more local and user-specific needs.

Private organizations have been involved in their own versions of health technology assessment for some time. In 1981, for example, the American College of Physicians (ACP) established a Clinical Efficacy Assessment Project to evaluate procedures, tests, and therapeutic interventions within the purview of internal medicine ( 16). Although primarily intending its guidelines to be used by physicians to eliminate obsolete and unnecessary tests and procedures, ACP also specified that the guidelines might be helpful to others for policymaking and for setting research agendas.

In 1982. the AMA established its Diagnostic and Therapeutic Technology Assessment Program (DATTA). Unlike the ACP effort, AMA's assessments were based on opinion surveys of selected panels of up to 70 physicians, and they were specifically aimed at assessing the acceptability and effectiveness of new technologies (20). Although the immediate purpose of the program was to provide information on technologies to physicians "in a timely manner," it was also a defense against the assessments of nonphysicians and was intended "to represent the views and concerns of the practicing medical community to public policy makers" (508). Both the ACP and AMA

[2] In 19X9. when NCHSR was folded into the newly created AHCPR, its full name was the National Center for Health Services Research and Health Care Technology Assessment.

## BOX 6-2: The Activities of AHCPR's Office of Health Technology Assessment, 1991-94

*Reviews* [1]

1991 Laparoscopic Cholecystectomy

1992 Home Uterine Monitoring
Procuren: A Platelet-Derived Wound Healing Formula
Cochlear Implantation in the Outpatient Setting

1993 Lymphedema Pumps: Pneumatic Compression Devices
Intradialytic Parenteral Nutrition for Hemodialysis Patients
Small Intestine and Combined Liver-Small Intestine Transplantation
External and Implantable Infusion Pumps

1994 Electrical Bone Growth Stimulation and Spinal Fusion

*Assessments*

*1991* Intermittent Positive Pressure Breathing Therapy
Hyperthermia in Conjunction with Cancer Chemotherapy
Cardiac Rehabilitation Programs
Polysomnography and Sleep Disorder Reports
Single and Double Lung Transplantation
Measuring Cardiac Output by Electrical Bioimpedance

*Forthcoming assessments*

Heart-Lung Transplantation
Plethysmography
Combined Kidney-Pancreas Transplantation

[1] In the terminology used by OHTA, 'Technology Reviews are brief evaluations of health technologies prepared by the Office of Health Technology Assessment, Agency for Health Care Policy and Research (OHTA/AHCPR) of the Public Health Service Reviews may be composed in lieu of a technology assessment because the medical or scientific questions are limited and do not warrant the resources required for a full assessment the available ewdence is limited and the published medical or scientific literature is insufficient in quality or quantity for an assessment. or the time frame available precludes utilization of the full. formal assessment process" (825)

SOURCE B Gordon Off Ice of Health Technology Assessment, Agency for Health Care Policy and Research Public Health Service, U S Department of Health and Human Services, Rockville, MD personal communication May 25 1994, U S Department of Health and Human Services, Public Health Ser'wee, Agency for Health Care Policy and Research Off Ice of Health Technology Assessment OHTA Assessment and Reviews Published 1981-, unpublished document Rockville MD February 1994

technology assessment activities still continue (see appendix C).

The extent of the blossoming of the private sector market is hard to evaluate precisely. In 1988 (he IOM attempted to document all U.S. producers of health technology assessments in its Medical Technology Assessment Directory (367). At that time IOM identified seven governmental and 30 nongovernmental organizations that performed assessments (table 6-1 ). Of the private-sector organizations, over half ( 16) were provider or payer organizations, such as the AMA, the

## TABLE 6-1: Organizations Involved in Health Technology Assessment Activities, 1988

| Type of organization | Name of organization |
|---|---|
| **Government** | U S Congress |
| | Prospective Payment Assessment Commission |
| | Off Ice of Technology Assessment |
| | U S Department of Health and Human Services |
| | Agency for Health Care Policy and Research |
| | Food and Drug Administration |
| | Health Care Financing Administration |
| | National Institutes of Health |
| | U S Department of Veterans Affairs |
| **Academic** | Brandeis University |
| | Health Policy Center, Organ Procurement Project |
| | Duke University |
| | Center for Health Policy Research and Education |
| | Georgetown University Medical Center |
| | Institute for Health Policy Analysis |
| | Harvard University |
| | School of Public Health, Institute for Health Research |
| | Johns Hopkins University |
| | Program for Medical Technology and Practice Assessment |
| | University of California, San Francisco |
| | Institute for Health Policy Studies |
| | University of Pennsylvania |
| | Leonard Davis Institute of Health Economics |
| **Provider/payer organization** | American Academy of Neurology |
| | American Academy of Ophthalmology |
| | American Academy of Pediatrics |
| | American College of Cardiology/American Heart Association |
| | Task Force on Assessment of Cardiovascular Procedures |
| | American College of Obstetricians and Gynecologists |
| | American College of Physicians |
| | American College of Radiology |
| | American Dental Association |
| | American Diabetes Association |
| | American Gastroenterological Association |
| | American Hospital Association |
| | American Medical Association |
| | American Society for Gastrointestinal Endoscopy |
| | Blue Cross and Blue Shield Association |
| | California Medical Association |
| | College of American Pathologists |
| **Other private** | Battelle Memorial Institute |
| | ECRI |
| | Lewin and Associates, Inc |
| | Medical Technology and Practice Patterns Institute |
| | Policy Analysis, Inc |
| | Project HOPE Center for Health Affairs |
| | U S. Administrators Inc |

SOURCE Institute of Medicine, Council of Health Care Technology *Medical Technology Assessment Directory* C Goodman (ed ) (Washington DC National Academy Press 1988)

ACP, and the American Hospital Association. Another seven were university-based centers (e.g., the Johns Hopkins University Program in Medical Technology and Practice Assessment), and seven were private consulting or research organizations (367) (table 6-1 ).

Since IOM's inventory, the organizations it described all still appear to exist, and at least some have grown considerably. Among the largest private firms is ECRI,[3] which in 1988 produced primarily technical reports on the capabilities of medical devices. It has doubled in size and has greatly expanded its breadth of assessments, producing about 40 assessments per year for clients ranging from providers to purchasers of health care ( 128,579). It also operates a technology assessment clearinghouse funded by the World Health Organization.

Entirely new firms have sprung up as well, capitalizing on the interest in effectiveness research, cost-effectiveness analysis, and assessments of individual technologies. Technology Assessment Group, Inc., for instance, incorporated in 1990, markets its expertise in cost-effectiveness analysis and quality-of-life studies (885). One novel company, MetaWorks, offers meta-analyses of clinical studies (652). An upcoming directory produced by ECRI will list over 200 organizations in the United States and elsewhere that undertake health technology assessments or related activities (579). The rapid growth of these activities attests to the increasing importance given to knowledge of the costs and health effects of specific medical technologies in private sector (and state-level) decisionmaking.

Activity in the private sector is especially interesting in light of the fact that it was opposition by manufacturers and health care providers that helped bring about the demise of NCHCT in 1981 (68,598). Ten years later, a collaborative group of manufacturers. payers, and providers in Minnesota has published a consensus document advocating technology assessment that is being used in State health reform efforts (329,5 18).

Growth is not confined to proprietary consulting firms. Hospitals and managed care providers are now entrenched consumers of technology assessment. Although relatively few individual hospitals conduct formal assessments (900), hospital organizations are producing them in significant numbers. The American Hospital Association, for example, issues a periodical *(Technology Reports)* that offers in-depth commentary on new technologies. The University Hospitals Consortium, an association of academic teaching hospitals, has had an in-house technology assessment office since 1989 (498). The Hospital Association of New York State recently produced a detailed manual for hospitals on how to do and use technology assessments for hospital decisionmaking (121 ).

Insurers have likewise begun to turn to formal technology assessments to assist their decisionmaking. In some of the most striking examples:

- Blue Cross and Blue Shield Association's Medical Necessity Program is now in its 18th year of operation. The Association has also expanded its Technology Evaluation and Coverage program through a cooperative technology assessment venture with Kaiser Permanence Medical Care Program, and in a major change from past policy the organization will make these assessments available to the public (282) (box 6-3).
- Other major insurers such as Cigna, Prudential, and Aetna now also have their own full y staffed technology assessment divisions (1 78).
- A managed care organization, The HMO Group, established its TEMINEX project in 1989 to assess technologies on behalf it its members (258).
- The Health Insurance Association of America, whose members tend to be somewhat smaller insurance companies, has investigated an

[3] ECRI (formerly the Emergency Care Research Institute) is now the full name of this organization.

**BOX 6-3: The Technology Assessment Activities of the Blue Cross and Blue Shield Association**

One of the earnest organized private technology assessment efforts was the Blue Cross and Blue Shield Associations (BCBSA's) Medical Necessity Program, which began in 1976 The Medical Necessity Program identified lists of medical and surgical procedures which contributed to the cost of health care but, in many Instances, did not make parallel contributions to the quality of care'" (67) The program's purpose was to inform member plans regarding specific coverage decisions and participation questions Physician organizations, such as the American College of Physicians, the American College of Radiology, and the American College of Surgeons, assisted in the identification of procedures that were either unproven, redundant when performed in conjunction with others, or repeated without clinical value The technology assessment process Included a literature review of articles and the creation of a guideline that was reviewed by the appropriate medical specialty society and the BCBSA Medical Advisory Panel Once approved, the guideline represented the official recommendation of the Blue Cross and Blue Shield Association to its member companies (67)

The Medical Necessity Program continues but has been augmented by a separate Technology Evaluation and Coverage Program, established in 1985 This programs goal is to assist member plans in determining the clinical status of emerging technologies and to aid in the coverage and reimbursement decisions The TEC Program evaluates medical and surgical procedures for specific conditions, focusing on the diagnostic and treatment value Unlike the Medical Necessity Program, it is explicitly concerned with costs as well as health effects Recently, Blue Cross and Blue Shield Association announced its decision to undertake its TEC efforts in collaboration with Kaiser Permanente Medical Care Program, a major prepaid care provider (282)

SOURCE Office of Technology Assessment 1994 based on sources as shown Full citations are at the end of the report

agreement with a private company to perform technology assessments for its members (329).

## Clinical Practice Guidelines and Technology Assessment

Clinical recommendations based on the deliberations of groups have become commonplace, as advances in medical knowledge have increased the complexity of decisonmaking and made it difficult for individual clinicians to keep abreast of the emerging literature. Many health professional associations themselves produce practice guidelines: AMA's *Directory of Practice Parameters* lists 1,500 guidelines of some kind produced by more than 45 organizations (22).

Clinical practice guidelines have a diverse array of potential roles and applications, reflected in the many definitions of guidelines that exist. Probably the most widely cited definition is one developed by the IOM. Here. practice guidelines are defined as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (371 ). This definition emphasizes the traditional role of guidelines in assisting in individual clinical decisionmaking.

Other definitions have emphasized the role of clinical practice guidelines as clinical policy statements about the proper way to practice clinical care. Woolf, for example, uses "practice guidelines" to refer to "the official statements or policies of major organizations and agencies on the proper indications for performing a procedure or treatment or the proper management for specific clinical problems" (944). Eddy distinguishes among different types of "practice policies, " which range from "standards" to "opinions, " ac-

cording to the level of certainty that underlies their respective recommendations (200).

Clinical practice guidelines clearly serve policy purposes other than establishing clinical policies and aiding in individual clinical decision - making. In its 1992 report on the topic, the IOM identified fivc major purposes of guidelines (371 ):

1. to assist clinical decisionmaking by patients and practitioners:
2. to educate individuals or groups;
3. to assess and assure the quality of care (e.g., by establishing clinical indicators for quality assurance programs);
4. to guide allocation of resources for health care (e.g., insurance payment decisions): and
5. to reduce the risk of legal liability for negligent care (e.g., through laws that restrict the liability of physicians who were following practice guidelines).

Clinicians have frequently viewed guidelines developed for some of these roles, particularly those associated with payment, with some suspicion. Guidelines promoted by insurers are regarded by physicians as less credible than guidelines promoted by the physicians' own organization (768). Much of the antagonism against the old NCHCT related to the agency charge to develop "'standards" for the use of particular technologies (68).

Guidelines less associated with payment have inspired less concern, even if those guidelines are sponsored by government agencies. In fact, federal agencies have sponsored the development of clinical practice guidelines for topics within their purviews for years. The Centers for Disease Control and Prevention (CDC), for example. has been promoting recommendations regarding vaccinations since the 1960s, in its role as protector of the public health. The National Heart, Lung, and Blood Institute's cholesterol, asthma, and blood pressure guidelines are likewise well known (663,854,856,857). Until recently, the National Cancer Institute published and disseminated screening recommendations for many cancers (361).[4]

When Congress established AHCPR in 1989, it created a new, separate, and very visible additional guidelines effort through its mandate that AHCPR establish a Forum for Quality and Effectiveness in Health Care to produce clinical practice guidelines. The theory was that the panels developing the guidelines would use the results of effectiveness research, augment these findings with their own expert judgment, and come up with templates for the best qua] it y medicine. AHCPR's guidelines and effectiveness research efforts were purposefully located outside of the Health care Financing Administration (HCFA) to enhance their acceptability to providcrs by minimizing their association with Medicare's cost control objectives (295).

The federal government dive into broad-spectrum clinical guidelines at AHCPR in 1989 undcrlined the split that had gradually been growing between the activities labeled "health technology assessment" and those involving clinical practice guidelines development. At AHCPR, the split is evidenced in three ways. First, the guidelines development office was established as an entity entirely separate from OHTA, with little apparent overlap in activities between the two. Second, in contrast to the technology-specific focus of OHTA's work, AHCPR's clinical practice guidelines focus on the broad sets of interventions used in the management of a particular clinical condition, rather than on individual technologies. Third, AHCPR "technology assessments" are staff-generated. while guidelines are developed by external expert groups sponsoed by the agency. The extent of the conceptual split between guidelines and technology assessment at the agency is demonstrated by the fact that staff in the guidelines office are quite insistent that clinical practice guidelines and technology assessment are entirely different activities (501 ).

[4] NCI now assembles and evaluates evidence regarding screening tests but does not make recommendations (see appendix C).

Yet much of the distinction between health technology assessment and clinical practice guidelines is artificial. Like other technology assessments, guidelines can focus either on a single technology (e.g., the acceptable applications of a particular procedure) or on a technology-related issue (e.g., alternative technologies for managing a particular medical condition). NIH's Consensus Development Program was explicitly designed as a method of assessing medical technologies that produced statements for clinicians about the appropriate use of those technologies. Both the ACP and AMA technology assessments efforts described above also share these characteristics.

And, like other technology assessments, guidelines can be used for policymaking, including payment and other resource allocation policies. In fact, all clinical practice guidelines represent resource allocation decisions on the part of the persons creating the guideline. Those decisions may be to underscore current practice—i.e., resources should best be allocated as they are at present-or to change resource allocation—i.e., in favor of different practices, which use different resources. RAND's expert panels examining the appropriateness of different indications for particular procedures, described in chapter 2, are particularly explicit attempts to create guidelines to influence the resource allocation associated with technologies they assess.

Thus, in the context of public policymaking, clinical practice guidelines can be considered a particular form of technology assessment, where the assessors are an expert panel and the audience comprises not only program decisionmakers but individual clinical decisionmakers as well. When guidelines are sponsored by the federal government, the different potential "roles" of guidelines are simply the mechanisms by which the government can attempt to influence the content of clinical care. The technologies examined in the guidelines may be individual products or procedures, or they maybe the sets of technologies used within a management strategy.

The government's goals in developing guidelines are presumably to improve the effectiveness and quality of care, constrain the costs of care, or achieve other social objectives (e.g., improve the equitability of access to care). One of the attractions of guidelines development as an assessment mechanism is the fact that it involves representatives of some of those affected by the guidelines through their inclusion in the expert group creating the guideline.

All guidelines are not equally valid or equally effective. The IOM has suggested some of the attributes of a guideline that it considers desirable. including reproducibility, applicability, and clarity (box 6-4). IOM's criteria do not address the implications of how costs are considered (or not considered) when creating guidelines. Their criteria also do not address the interactions of the expert group and how group members consider the information available to them, another important contributor to the validity and reliability y of guidelines. These and other components of guidelines development are discussed in chapter 7. Chapter 8, in turn, discusses the impact of different strategies for implementing guidelines on clinical practice.

## CONCLUSIONS

**One of the most remarkable developments in the field of health technology assessment has been its transition from the public to the private sector.** Certainly, a few individual private sector payers and providers have been involved in health technology assessment for years. What is new is the degree to which technology assessments are becoming a standard ingredient in private-sector decisionmaking. While the federal government's investment in individual technology assessments has been largely unchanged in degree over the past decade, the private market in technology assessments has become a full-fledged economic activity in its own right.

Two seemingly opposing trends in this market are notable. The first is the increasing number of payers and providers, or groups of providers, performing their own technology assessments and with staff dedicated to that purpose. This trend is illustrated, for example, in Aetna's dedicated

## BOX 6-4: The Institute of Medicine's Attributes of a "Good" Guideline

The Institute of Medicine has proposed several attributes that a "good" guideline should have.

- *Validity* — when followed, practice guidelines should lead to the health and cost outcomes projected for them,
- *Reliability* — given the same evidence and methods for guidelines development, another set of experts should produce essentially the same statements and given the same clinical circumstances, the guideline should be interpreted and applied consistently by practitioners.
- Clinical *applicability* — practice guidelines should be as inclusive of appropriately defined patient populations as evidence and expert Judgment permit, and they should explicitly state the population to which statements apply,
- *Clinical flexibility* — practice guidelines should identify the specifically known or generally expected exceptions to their recommendations and discuss how patient preferences are to be Identified and considered.
- *Comprehensiveness* — practice guidelines should include all likely clinical alternatives or indications for the use of an Intervention.
- *Specificity* — guidelines should have detailed descriptions of the circumstances for which an intervention IS recommended, is appropriate, or for which there is inadequate information to form an opinion.
- *Soundness* — guideline recommendations must be based on good evidence
- *Ease of use* — guidelines should be concise, unambiguous, and in a format which makes it easy for clinicians to use them,
- *Scheduled review* — guidelines should include a statement about when they should be reviewed for revisions,
- *Documentation* — the procedures followed in developing guidelines, the participants involved, the evidence used, the assumptions and rationales accepted, and the analytic methods employed must be documented and described meticulously,

SOURCE Off Ice of Technology Assessment, 1994, based on information from Institute of Medicine, *Guidelines for C/mica/Practice from Development to Use* (Washington DC National Academy Press, 1992), L L Leape, "Practice Guidelines and Standards An Overview, " Qua/@ *Review Bulletin*, 16(2) 42-49 1990

technology assessment section and the creation of a technology assessment shop in the University Hospital Consortium. Dedicated in-house divisions such as these enable the organizations to develop technology assessments specifically tailored to the needs of their users—in these cases, an insurer and academic medical centers.

At the same time, consulting firms and academic centers specializing in technology assessments are flourishing. Rather than tailoring their assessments exclusively for the interests of one particular user, these organizations market a relatively more uniform product to multiple users. (Individual assessments, however, may be tailored for a specific client.) What both these trends-dedicated in-house technology assessment and greater use of external assessors—have in common is

their demonstration of the enormous and growing demand for assessments.

Health technology assessments generally require multiple areas of expertise (e.g., clinical, statistical, economic, etc.). While only relatively large organizations can justify many staff dedicated to the endeavor, private firms have responded to the market demand for health technology assessments by assembling the needed expertise in consulting firms and marketing that expertise to organizations that cannot sustain in-house efforts. The recent changes in the Blue Cross and Blue Shield Association's dedicated technology assessment division illustrate this nicely: the organization is now collaborating jointly with Kaiser Permanence in this effort, and it is marketing its assessments for the first time to outside organizations.

OHTA has been instructed by Congress to set priorities for technologies to assess in the event it can conduct some private assessments (Public Law 102-410), and it has taken steps to establish these priorities (827). Given the vastly expanded private sector capability for individual technology assessments, however, payers, providers, and others wanting assessments of particular technologies are not dependent on the government to obtain them. Thus, rather than expanding its activities to cater to the private market, one possible future role for OHTA would be to continue to perform assessments for government programs only. The Office could, however, also expand its usefulness to other government decisionmakers (e.g., Medicaid programs). 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 cannot be adequately conducted, in the private sector.

Alternatively, Congress may consider that developments in health reform underscore the need for reliable assessments from a single source. so that private payers and providers are not faced with conflicting conclusions from assessments by different sources, and so that critiques of the assessments can be both public and focused. If this is the case, OHTA (or another federal body) would need to greatly increase its size and scope to accommodate user needs.

Even if a more limited role is envisioned for OHTA, its usefulness might be improved by encouraging it to assess technologies with greater impact. Many of its past assessments have been on fairly technical and esoteric topics (e.g., the Rehfuss test for gastric acidity and the debridement of mycotic toenails). By broadening the scope of its assessments (and staff expertise) to include cost and other impacts, and extend the breadth of technologies it assesses, OHTA would be more likely to be able to help fill future needs under health reform.

In both its legislative origins and its organizational placement, the new federal guidelines effort is much more closely aligned with effectiveness research than with health technology assessment. At present, AHCPR guidelines tend to be viewed as distinct from technology assessments by virtue of their focus (management- vs. technology-focused); their purpose and audience (educational advice to clinicians vs. coverage decisions for payers); and their source of production ('*expert group" vs. staff-produced).

In fact, however, federal guidelines development efforts arc 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. From the perspective of public policymaking, the distinction between guidelines and technology assessments is not a valid one.

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. Because guidelines are important to many of the proposals to improve the health care system, in both the private and the public sectors, the methods by which they are derived and the impact they have on practice deserve considerable attention,

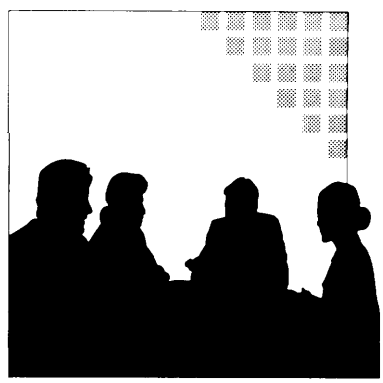
# The Development of Clinical Practice Guidelines

7

Clinical practice guidelines are increasingly being viewed as promising tools for promoting cost-effective and appropriate care (201-206,944). Of particular interest are guidelines that are based on the recommendations of panels of experts or representatives sponsored by various organizations. At least 1,500 such guidelines exist, issued by groups as diverse as physicians' professional associations, health care insurers, and the federal government (628).

The focus of this chapter is on federal activities, with a special interest in the efforts of Agency for Health Care Policy and Research (AHCPR) because it represents the latest guideline effort and one of particular interest to Congress. Selected, well-established private guideline efforts are also discussed here to put federal activities in a broader context. Chapter 8 looks beyond the guideline development methods and assesses the potential for guidelines to change clinical practice.

Six federal and four private-sector guideline development efforts form the basis for discussion here. Detailed descriptions of each guideline effort are presented in appendix C.



## FEDERAL GUIDELINE EFFORTS

1. **Agency for Health Care Policy and Research.** Since it was established in 1989, AHCPR has issued 11 guidelines, with another 10 under development. Guidelines are produced by panels of 15 to 18 members that are notable for their emphasis on including both consumer representatives and nonphysicians, as well as physicians from a variety of disciplines. AHCPR's guidelines generally address the clinical management of broad health conditions, such as cancer pain and heart failure, and take up to three and a half years to complete.

2. **NIH Office of Medical Applications of Research.** National Institutes of Health's OMAR has issued over 100 Consensus Development Statements since the program's inception in 1977. The primary mission of the Consensus Development Program is to identify and then disseminate to clinicians clinically relevant findings emerging from NIH research, and most topics for conferences are suggested by Institutes of NIH. OMAR's process is unusual for its brevity and its format. Although panel members receive some background information, the recommendations are developed over the course of a single, three-day Consensus Development Conference that includes substantial public input.
3. NIH National Heart Lung and Blood Institute. NHLBI has sponsored detailed guidelines on three medical conditions: high blood pressure, high cholesterol, and asthma. Unique to NHLBI's effort is its guideline panel structure; the guidelines are issued by very large panels (20 to 50 members) that are overseen by coordinating committees made up of representatives of professional societies, voluntary health agencies, and consumer organizations. The coordinating committees have an educational focus; they help promote the guidelines as well as perform other educational functions.
4. **NIH National Cancer Institute.** NCI has previously produced a number of guidelines on cancer prevention and management, but recently it has decided not to make explicit recommendations at all (305). Instead, NCI now issues evidence-based informational statements through its computerized PDQ (Physician Data Query) database. Standing "editorial" panels, which include both NCI staff and outside experts, review and interpret the literature and periodically update the statements on the database.
5. **CDC Advisory Committee on Immunization Practices.** The ACIP, probably the best known of the many groups within the Centers for Disease Control and Prevention (CDC) that issue clinical practice guidelines, comprises a 12-member standing committee that makes recommendations regarding immunization doses, schedules, and other issues with input from liaison representatives from professional societies and other federal agencies. Unlike most other federally sponsored guidelines, those of the ACIP generally are formally endorsed as government policy.
6. **U.S. Preventive Services Task Force.** The USPSTF, convened by the Office of Disease Prevention and Health Promotion (ODPHP), [1] was the first federally sponsored guidelines panel to rate the quality of the scientific evidence behind its recommendations and to link its recommendations directly to that evidence. It is unusual in that it limits group judgment to interpreting the evidence; the personal opinions of panel members are not considered relevant to the guidelines. The first Guide to Clinical Preventive Services, published in 1989, reviewed evidence of the effectiveness of 169 preventive services. It is now being updated and augmented.

## PRIVATE EFFORTS

1. **American College of Physicians.** Since 1981, ACP has developed more than 160 guidelines through its Clinical Efficacy Assessment Project (CEAP). Its guideline recommendations, like those of USPSTF, are rated according to the level of evidence supporting them, although the panels do not exclude a role for expert opinion. CEAP panels comprise only internists (the membership of ACP). Their process is unusual for its heavy reliance on consultant-produced reviews of the evidence as the basis for guidelines.
2. **AMA Diagnostic and Therapeutic Technology Assessment Program.** The American

[1] ODPHP is located within the Department of Health and Human Services, under the Assistant Secretary for Health.

Medical Association's DATTA program, in existence since 1982, uses an expert panel to provide forma] guidance regarding the safety and effectiveness of individual technologies (e. g., lung transplantation, Teflon™ preparations for urinary incontinence). Unlike other efforts that produce clinical guidelines. the DATTA process relies primarily on a mailed survey of the opinions of an expert panel; there is no interaction among the panel members.

3. Harvard Community Health Plan (HCHP). Practice guidelines in the form of clinical algorithms—structured flowcharts of decision steps and preferred clinical management pathways (see box 7-1 )—are developed as part of this health maintenance organization's quality improvement program. As of early 1994, over 30 clinical topics had been completed or were under development (e.g., asthma, colon cancer screening, depression ), most created by an internal panel of HCHP clinicians. Unlike most other guidelines efforts reviewed here, HCHP panels specifically consider cost-effectiveness during the algorithm development process.
4. **RAND Corp.** RAND has developed a method for using forma] group processes to rate the appropriateness of indications for medical and surgical procedures (e. g., hysterectomy. coronary angiography). The ratings have been used both retrospectivcly, to assess the appropriateness of care, and prospectively in precertification programs. The process includes nine-member multispecialty clinical panels that review background material on the literature and rate each possible indication for a procedure on a 9-point appropriateness scale, using a highly structured process of group interaction. Median ratings are used to describe the group judgments, and levels of agreement and disagreement arc formally defined.

## GUIDELINE ISSUES RELATED TO DEVELOPMENT

### Overview

The diverse federal and private efforts to develop clinical practice guidelines, discussed in this chapter and described in greater detail in appendix C, share a number of features. Most groups devel - oping guidelines have in common the objective of improving clinical decisionmaking in some way by providing clinicians, and sometimes the public. with information. All assign the basic task of creating or endorsing the guideline recommendations to a panel of appointed experts or representatives. In the case of guidelines issued by federal agencies, the guideline panels are virtually always groups of external advisors; most agencies issue, but do not formally endorse, the guidelines created by these groups.

All of the guideline efforts also have some process for identifying potential guidelinc topics, for extracting relevant background information from the scientific literature to which panel members can refer, and for eliciting judgments and (usually) additional opinions or experiences about the literature and the topics from panel members. Most also convene panel members in person to discuss recommendations. Guidelines are usually issued as a book, an article, or statement that includes recommendations to clinicians regarding clinical practice.

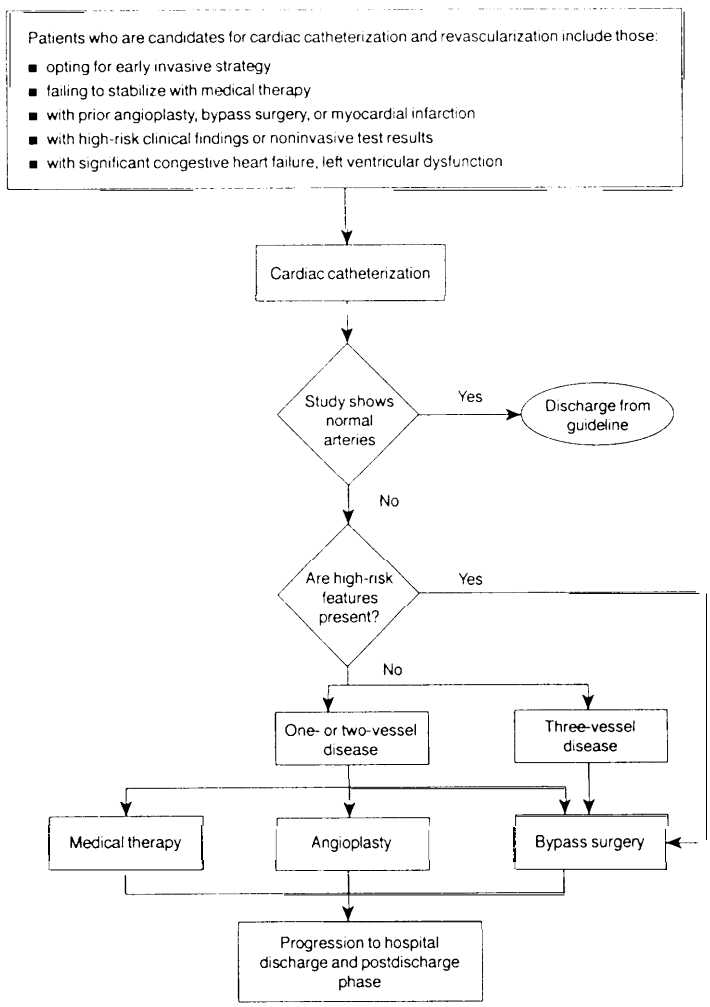
Despite the similarity in the basic structure of guideline development activities. the actual methods of the different groups vary considerably. Major features of guideline development that tend to distinguish one approach from another am:

- **The way in which guideline topics are selected.** Some agencies (e.g., AHCPR) have statutory direct ion regarding guidelines topics. Others (e.g., ACP and AMA's DATTA program) generate topics internally by various means, while still others (e. g., OMAR and RAND) primarily generate guidelines on topics proposed or endorsed by external sponsors.
- **The characteristics of guideline panels and the processes and criteria for selecting panel members.** Guideline panels usually include between 10 and 20 individuals. Some are homogeneous, including only members of a particular group (e.g., a professional society), while others include a range of individuals such as health care providers, methodologists, and

## BOX 7-1: Presenting Guideline Recommendations Through Clinical Algorithms

Algorithms are powerful tools for making explicit the relationship between clinical states and diagnostic and therapeutic decisions where there is diagnostic certainty (e.g., if positive strep test, then antibiotic therapy) or diagnostic uncertainty with a probably benign outcome (e.g., if probably viral throat infection, then culture and wait to treat). Algorithms enable the clinician to practice a defined standard of care and may be translated into protocols or chart audits (490)

The Harvard Community Health Plan develops clinical algorithms as part of its quality assurance program. Clinicians seem to prefer algorithms over prose descriptions of the decisionmaking process (490). The National Heart Lung and Blood Institute Illustrates its recommendations with algorithms, and some of the Agency for Health Care Policy and Research panels have also used algorithms to illustrate their recommendations and to identify patient counseling and decision points (figure 7-2) (321). An example of AHCPR's algorithm for management of patients undergoing cardiac catheterization is described below.




SOURCE Off Ice of Technology Assessment, 1994, based on sources as shown Full citations are at the end of this report

consumers. Few organizations sponsoring guidelines. however, have detailed, documented rules regarding panel composition.

- **The scope and perspective of the guidelines.** Some guidelines consider the relative benefits and harms of a wide variety of the alternative clinical approaches for a particular condition or complaint (e. g., AHCPR and NHLBI), while others target particular procedures or technologies and describe their appropriate uses (e.g., DATTA and RAND). Almost all guideline panels consider the safety and effectiveness of interventions, but increasingly, guideline panels are addressing broader issues such as cost-effectiveness, patient preferences. and aspects of health system organization that affect the use of the interventions under consideration. Guidelines also differ in whether they are targeted to specialists, primary care providcrs, or other potential users.
- **The processes used to extract evidence and other information from the scientific literature, experts, the public, and other sources. Some** guideline processes emphasize exhaustive literature searches and syntheses, while others work without a formal analysis of the quality of evidence available to them, or descriptive information on the current state of medical practice.
- **The group processes used to consider evidence and produce agreement on recommendations.** Many panels have fairly loose, free-flowing discussions through which they debate evidence, opinions. and recommendations. Others, however. emphasize formal ways to structure the interaction and judgments of panel members.
- **The degree to which the methods used by panels are explicit, documented, and available.** The processes of some guideline groups are described in great detail within the guidelines themselves and in professional journals (e.g.. HCHP, RAND), Other groups have not published any descriptions of how their guidelines were developed (e.g., CDC's ACIP. NHLBI).
- **The extent to which guideline recommendations are linked directly to scientific evidence.** Some, guideline panels rely primarily on scientific evidence as the basis for recommendations (e.g.. USPSTF and ACP), while others rely on the opinions and judgment of experts to make recommendations when evidence is lacking (OMAR, RAND, AHCPR). In at least one case, prescriptive recommendations are no longer made at all; NC I recently decided to provide informational statements to physicians, which simply interpret existing evidence. rather than specific recommendations for practice.
- Administrative features of the process, such as whether guideline panels are "standing" or ad hoc and the extent of administrative oversight of guideline activities.

These features are discussed in more detail below.

## Choosing Guideline Topics

Many organizations that are developing guidelines publish the criteria and process they use to select topics. Criteria frequently cited as being used to select topics for guideline development include:

- **Public health impact—the** prevalence, incidence, and severity of the condition in question and the potential for interventions to prevent the condition or ameliorate symptoms.
- **Cost of procedure—a** procedure might be costly as a single unit (e. g., organ transplantation) or because it is commonly performed, for example, as part of population screening (e.g.. colonoscopy).
- **Availability of evidence—for some** technologies **there is** good evidence on which to base judgments (e.g.. several randomized clinical trials), while for others only descriptive clinical experience and opinion are available.
- **Variation in clinical practice—may** reflect clinician uncertainty or genuine differences in schools of thought in the rnanagement of certain conditions.

- **Controversy —may be over alternative interventions for a condition, who should deliver care. or where a service should be delivered.**
- **New versus established technologies+ stab**lishing guidelines on a new or emerging technology could forestall inappropriate use.

In general, however, these criteria serve more as loose guides than as part of a systematic prioritization process. For example, many groups select topics based on the level of controversy and availability of evidence, but most do not try to assess the state of clinical practice or the quality of evidence for a particular guideline topic *before* a guideline topic is selected. Instead, a guideline topic is selected usually through some sort of nomination or survey process, and then a panel is assembled to focus the assessment and begin to identify relevant evidence.

Federal agencies often have congressional mandates that give some direction to their selection of topics. AHCPR's guidelines effort, for example, is specified by its authorizing statute, which directed the agency to examine issues of relevance to the Medicare and Medicaid populations (Public Law 101 -239). When reauthorized in 1992, AHCPR was further directed to consider clinical treatments or conditions that were costly, for which there was significant variation in the frequency or the kind of treatment provided, and for which inappropriate use of health care resources was likely (Public Law 102-410).

The authorizing legislation also specified that AHCPR, created in December 1989, had to issue at least three guidelines by January 1991 (Public Law 101-239). To reach this deadline, AHCPR initially selected topics for which guideline development was already underway (798). Since 1993, AHCPR has published a list of possible topics for guideline development in the *Federal Register* and elsewhere and solicits comments and recommendations for new topics (812). AHCPR has also recently brought representatives of groups together to discuss potential topics for guidelines (53). The Institute of Medicine (IOM) is currently conducting a study for AHCPR on setting priorities for guideline development (81 3).

Guideline efforts within NIH emphasize the role of disseminating research findings of the Institutes to clinicians. Topics for OMAR's Consensus Development Conferences are suggested by the Institutes themselves. In addition, two of the institutes, NHLBI and NCI, issue their own guidelines or statements on topics within their domains. NHLBI focuses on only a few clinical conditions that fall within its purview (i.e., high blood pressure, cholesterol, and asthma), while NCI issues statements on topics across the spectrum of cancer management (i.e., screening, treatment, and supportive care).

AHCPR, OMAR, NHLBI, and NCI all frequently cover preventive services, such as screening and immunizations, in their guidelines. In addition, two other federal guideline efforts reviewed here—CDC's ACIP and ODPHP's USPSTF—focus exclusively on preventive practices. USPSTF covers the full range of preventive services provided in clinical settings, while ACIP makes recommendations relating to immunization practices.[2]

Some private guideline sponsors have developed more systematic ways to solicit opinions on potential topics for guidelines from practicing clinicians. ACP, for example, surveys its members to help identify topics of interest to practicing internists as part of its CEAP program. Topics are selected for AMA's DATTA program in part through a survey of DATTA subscribers.[3] HCHP develops guidelines based on nominations made by HCHP clinicians and also considers health plan data to identify practices for which there is variation.

[2]CDC also issues other guidelines. Topics generally focus on achiev ing national health objectives as stated in *Healthy People 2000* and in CDC's mission statements (709).

[3]Questions for DATTA evaluations are considered from a variety of sources (e.g., physicians, patients, third-party payers, peer review ers). The survey is one of several mechanisms used to identify topics (see appendix A).

Organizations vary greatly in the type of topic selected for guideline development. Some focus on selected conditions and complaints while others address specific procedures or technologies (945). AHCPR has assumed an interdisciplinary perspective to examine alternative approaches to diagnose and manage selected chronic conditions and complaints (e.g., pressure ulcers, pain. depression). For example, they examined medical, surgical, and behavioral interventions in their urinary incontinence guideline. Such an approach is attractive to primary care providers and patients in that it provides an assessment of the relative value of competing approaches that may never before have been compared with one another in a single document. These comprehensive and interdisciplinary guidelines are generally more demanding to produce from a methodologic point of view (e.g., nomenclature, measurements, and outcomes of interest may differ across disciplines) and may therefore take more time and be more costly than a more narrowly focused guideline. Some of the AHCPR guidelines, for example, have taken over three years to develop.

Other efforts are more narrowly focused on the circumstances for which particular technologies are most appropriately used. RAND researchers have developed appropriateness ratings for expensive, commonly used surgical procedures (e.g., coronary angiography, hysterectomy ).4 OMAR has tended to focus on technologies emerging from the NIH research arena and in (he process of diffusing into clinical practice (e.g., using antibiotics to treat peptic ulcers). The AMA DATTA program evaluates primarily new and emerging technologies of interest to specialists (e.g., lung transplantation). These more narrowly focused assessments have their attractions: they can target a technology about which there is uncertainty within the practice community. they can be targeted to certain groups of clinicians, they are attractive to insurers and health planners, and they can often be completed relatively quickly. Most OMAR and DATTA assessments, for example, are completed within one year.

## Selecting Guideline Panels

The diversity of guideline efforts is reflected in the composition of guideline panels, which vary in size and include a range of individuals from technical experts to consumers (609). Most guideline panels range in size from 10 to 20 members. NHLBI panels are unusually large, including as many as 50 members, with smallcr subcommittees formed to address specific subtopics.

The background and perspectives of the individuals involved in guideline setting likewise vary considerably, across both different guideline efforts and different panels within a particular effort. Federally sponsored guideline groups are often relatively diverse, including nonphysicians and consumer representatives. The AHCPR and NHLBI panels, for example, usually include a range of health care providers and at least one consumer representative. In contrast, private physician groups have generally confined panel membership to physicians. The ACP CEAP and AMA DATTA panels, for example, include only physicians, some of whom have methodologic expertise.

A potential threat to the validity of a guideline is selecting panel members who share a particular bias. A biased group could be assembled quite inadvertently by selecting certain types of members. In research on panels using the RAND method. all-surgical panels rated more procedures appropriate and had more agreement about appropriateness than a mixed panel composed of surgeons and non surgeons (448,683). Within mixed panels. surgeons rate the appropriateness of surgical procedures substantially higher than do nonsurgeons (585). This finding is consistent with others showing that physicians who perform a given interven -

[4] More recently RAND has looked at less invasive procedures (e.g., spinal manipulation for low-back pain) (689).

(ion frequently are more likely to judge it as beneficial (578).

This user bias is not surprising, as one would expect those who perform a procedure to be more committed to its value. What is striking is the magnitude of the effect. In one example, a panel of surgeons assessing carotid endarterectomy rated 70 percent of cases as appropriate, whereas a multidisciplinary group found only 38 percent of procedures to be appropriate (448).[5]

User bias is not the only source of panel difference; different backgrounds and cultural assumptions also matter. Panels in the United States and the United Kingdom (with the same physician specialty composition) came to different conclusions when assessing the appropriateness of treatments for coronary disease. The U.S. panel judged more indications appropriate, and had better agreement among members, than did the U.K. panel. When the ratings of the two panels were applied to two groups of patients who had had the procedure in question, the U.S. panel judged 17 percent and 27 percent of the procedures as inappropriate, whereas 42 percent and 60 percent were judged as inappropriate by the U.K. panel counterparts (85).

Panel sponsorship, composition, and the inherent interests of different groups of clinicians can be major factors in the acceptance of guidelines. ( 138,582). Some eye surgeons, for example, disagreed with some aspects of AHCPR's cataract guideline. The AHCPR panel reviewed evidence on the criteria used to determine when cataract surgery might be indicated and found no evidence to support the use of some of the preoperative tests now in use. The AHCPR panel recommended that a patient's level of visual dysfunction rather than certain other tests be used as a criterion for surgery. The surgeons contended that the federal guidelines were intended to reduce the number of Medicare patients who would be eligible for cataract surgery. In another example, the American Psychological Association, the principle professional society representing psychologists, failed to endorse AHCPR's depression guideline, in part because of its perceived emphasis on medical therapy at the expense of psychotherapy (618).

To enhance guideline credibility, most clinical practice guideline efforts have panels drawn heavily from clinician groups whose practices will be affected by the guideline. AHCPR and NHLBI, for example, solicit nominations for panel membership from health care professional organizations. The guidelines developed at HCHP are written by the very clinicians who will ultimately use them. (The research suggesting that clinicians are more likely to believe guidelines in whose development they participated is discussed further in chapter 8,)

The desire to appoint guideline panels that are credible to the clinicians whose practices will be the most affected presents a dilemma for policymakers, because it also creates the potential for biased guidelines when developed by enthusiasts. Guideline panels that are intended to represent affected clinicians are likely to comprise a disproportionate number of users.

Another limitation of homogeneous clinician panels, particularly panels comprising primarily physicians, is the inability of such a panel to represent nonphysician concerns. It may be easier to consider interventions outside of the usual purview of medical specialists, and issues such as patient preferences and concerns, with a more heterogeneous panel. The inclusion of nurses and a psychologist on AHCPR's urinary incontinence panel, for example, probably facilitated the panel's consideration of interventions such as biofeedback techniques. An important recommendation of the urinary incontinence panel was

[5] The RAND process rates indications as appropriate, equivocal, or inappropriate. The proportion of *inappropriate* cases fell from 31 percent to 19 percent when a multidisciplinary instead of an all-surgical group rated the procedures. Of note is that both panels rated the same 12 to 13 percent of cases as inappropriate. This indicates that there appears to be a consensus regarding inappropriateness for a small subset of cases (448).

that such techniques be considered seriously as treatment options that could be alternatives to surgery. which has inherent risks and complications (802). Heterogenous panels too have their limitations, however. While consumer representatives and panel members representing fields such as ethics may play important roles in setting the guideline agenda and expressing the possible personal and social consequences of guideline recommendations, they often lack a technical background and so may not be able to fully participate in panel deliberations regarding the interpretation of medical and epidemiological evidence under consideration.

Despite the importance of panel membership on the scope and recommendations of a guideline, few organizations have strict criteria or rules regarding panel membership. Government-sponsored panels on occasion have been accused of bias for including enthusiasts for particular interventions (312). Of the processes reviewed for this report, only RAND and AMA's DATTA set a limit for the number of panelists that perform the procedure under consideration. AHCPR's reauthorizing legislation required that panel members who derive their primary income from procedures under consideration be limited on the panel but did not specify what the limit should be. Some groups attempt to screen panelists for potential conflicts of interest (e.g., AHCPR, NIH, ACP). OMAR seeks a chairperson and panelists who are neutral (388).To try to assure neutrality, the publications of candidate panelists are scrutinized to ensure that they have not published extensively on the conference topic (378).

## Defining the Scope of Guidelines

The intended audience for a guideline is an important determinant of the guideline's scope. Guidelines are typically directed at physicians, but issues of importance to other health care providers (e.g.. nurses, chiropractors), patients, and payers have broadened the focus of some guideline efforts. Federal guideline efforts have generally focused on primary care clinicians and increasingly have been directed to patients. AHCPR, NHLBI, and NCI, for example, direct their guidelines and statements to both clinicians and patients.

Safety and effectiveness are issues addressed in almost all guideline efforts. Given that guidelines are policy statements about the appropriate distribution of clinical resources, however (see chapter 6). some observers argue that unrealistic or even undesirable recommendations can be made when factors such as cost, health care system constraints. and patient preferences are not considered in the process of examining alternative clinical approaches. For example. a recent recommendation of the NIH Consensus Development Program-that all infants be screened for hearing impairment within the first three months of life (preferably before discharge from the hospital)—has been criticized, in part because many practical implementation and cost issues were not fully addressed (59,1 46). Similarly, some argue that implementing the NIH Consensus Panel recommendation that primary care physicians refer elderly patients suffering from sleep disorders to centers for sophisticated testing would be prohibitively expensive, because the condition is very prevalent (701 ). While some policy makers argue that guideline developers need to consider the health policy implications of their guidelines while they are being developed, scientists involved in guideline development have often expressed their discomfort in assuming this role more explicitly (633 ).

### *Cost and Cost-Effectiveness*

The role of cost projections and cost-effectiveness analysis in developing practice guidelines is controversial (945 ). Many federal guidelines have included assessments of the guidelines' likely impact on health care costs, and have included some informal discussion of existing evidence of cost-effectiveness, but none reviewed by OTA for this report has routinely included formal cost-effectiveness analyses in the recommendation making process. Increasingly, guideline developers have included resource assessments in their guidelines,

but groups differ in how much and how explicitly they allow costs to influence their recommendations.

Most guidelines issued by AHCPR to date have included statements about some of the anticipated changes in health spending that would associated with guideline implementation, but they have not explicitly considered the relative cost-effectiveness of alternative interventions when making guideline recommendations. Nor, often, have the anticipated savings from implementing an AHCPR guideline been compared directly and quantitatively with new costs that the guideline would impose (e. g., by encouraging the use of certain services). The recently released AHCPR guideline on heart failure, for example, was promoted with a discussion of the fact that its implementation could result in savings of $2 billion per year (448a). The guideline as promoted, however, did not present quantitative estimates of offsetting new costs. A few AHCPR panels have not considered cost explicitly at all (e.g., the guideline on management of HIV infection).

HCHP guidelines panels do frequently and explicitly consider costs and cost-effectiveness in their deliberations. NHLBI has added some discussion of costs in its most recent guideline on managing high cholesterol (857). The USPSTF, NCI, and RAND panels sometimes review evidence of cost-effectiveness, but they have explicitly excluded cost as a criterion for their recommendations or judgments regarding appropriateness.

Even when panels do attempt to incorporate cost considerations in guideline development, cost data are often not available for all interventions under consideration. AHCPR has commissioned a study on sources of cost data for guideline development, and the agency reports that it is assessing the adequacy of the cost analyses included in 10 of the guidelines it has sponsored (821).

The IOM has concluded that every clinical guideline should include information on the health and cost implications of alternative management strategies, but that every guideline need not be based on formal judgments of cost-effectiveness. They reasoned that this charge may be too great for individual guidelines panels and that perhaps guideline developers were not always the right source of such judgments (376). The Institute did not explore in detail exactly how cost-effectiveness considerations should be integrated into guideline development, however.

### *Patient Health Status and Functioning*

There is great interest in formulating clinical recommendations based on health status assessments that are of interest and relevance to patients (376). Measures such as maintenance of physical, cognitive, and social functions and alleviation of pain and discomfort are especially relevant when developing guidelines for non-life-threatening chronic illnesses. NCI, for instance, maintains information on supportive cancer care (e.g., managing cancer-related nausea and pain) as part of its computerized PDQ database for patients and clinicians (359). While functional outcomes are widely regarded as important, information on them is often unavailable because many clinical studies have not included them as outcome indicators.

Sometimes functional patient outcomes are considered, but there are limits to how accurately they can be assessed. All guideline panels examined in this report comprise primarily clinicians, and many studies of patient outcomes are based on clinicians' assessments. As discussed in chapter 3, however, clinicians' and patients' assessments of outcomes and their importance can be very different. Nor do clinicians' value judgments, used throughout the guidelines process, necessarily reflect patient and societal values (884).

It is also unclear how much weight the outcomes are given. When clinicians rate appropriateness as part of the RAND process, for example, a variety of patient outcomes may be considered. RAND general] y defines appropriateness to mean that "the expected health benefit (increased life expectancy, relief of symptoms, reduction in anxiety. improved functional capacity, etc. ) exceeds the expected health risks (mortality. morbidity, pain produced by the procedure)" (688). What

mix of outcomes physicians use in their assessments and what relative weights these outcomes are assigned is unknown. The meaning of such assessments is even open to question, because some evidence suggests that physicians are often poor judges of levels of patients' discomfort and functional status.

### *Patient Preferences*

Patient preferences are measures of sat is faction or desirability that people associate with the presence of symptoms and functional limitations that can affect quality of life (268). Incorporating patient preferences into guidelines is of great interest, but how to measure and usc patient preferences are subjects of ongoing research and debate. The quantification of patient preferences (also called patient utilities) is an active area of research, but there have been few attempts to formally integrate such patient preferences in the guideline process. A notable exception was Oregon explicit incorporation of patient preferences when initially prioritizing health services to establish a benefit package under its controversial Medicaid reform plan (788).

AHCPR and NHLBI have included patient representatives on guideline panels. and AHCPR and the NIH Consensus Development Program routinely hold public forums as part of the guideline process where public concerns and questions can be aired. AHCPR has pioneered a very pragmatic way to include patient preferences into their guidelines. Clinical algorithms are used to portray recommended management strategies, and the algorithms include points in the decisionmaking process where physicians and other caregivers need to discuss with patients or families their preferences for particular options (32 1,376). Assessments of patient preferences were made in the AHCPR guideline on benign prostatic hyperplasia. Wide variations in preferences were found, leading the panel to conclude that patients preferences need to be elicited as part of the treatment decisionmaking process (819).

Recognizing patient preferences in electing treatment options is clearly desirable for patients, but it does not necessarily lead to decreases in observed variation or more standardized practice. In one study, patients who had experienced acute upper gastrointestinal bleeding almost always preferred to have diagnostic endoscopy rather than less invasive tests or no testing. because they found the information it conveyed comforting even though it would not affect their management or prognosis. The researchers concluded that the current rate of diagnostic endoscopy is higher than would be expected based on physicians preferences but is quite consistent with patient preferences ( 1 77).

## Identifying and Synthesizing Evidence

All of the guideline efforts reviewed here include some mechanism for identifying and synthesizing the existing literature relevant to the guideline topic, so that it can be considered and discussed by panelists. The way in which this is carried out, however. varies considerably.

In many cases, extensive literature reviews are conducted as part of the guideline development process, sometimes at great expense. At AHCPR, for example, literature reviews have taken up to nine months and have cost up to $235,000 (376).

Once the evidence is amassed, different strategies can be used to synthesize it. Often panel members assess the literature themselves. The NHLBI guideline process, for instance, leaves literature reviews to panel subcommittees with no set methods or criteria established to ensure uniformity within a guideline.

Panels sometimes are assisted by a methodologist trained in epidemiology or statistics. AHCPR panels, for example. have generally benefited from the assistance of methodologists assigned to the panel to construct evidence tables. A number of analytic techniques have been developed to synthesize clinical evidence (e.g., meta-analysis). but the techniques are not usually used because they arc time consuming and differences in study characteristics often preclude their use.

A few guideline panels actual] y rate the quality of available evidence and give the most weight to high-quality studies (e.g., AHCPR, NCI.

**BOX 7-2: U.S. Preventive Services Task Force Strength of Recommendations, Design Categories, and General Criteria of Effectiveness**

**Strength of recommendations**

- A There is good evidence to support performing the preventive service,
- B: There is fair evidence to support performing the preventive service
- C There is poor evidence to support performing the preventive service, but recommendations may be made on other grounds.
- D There is fair evidence to discontinue performing the preventive service
- E, There is good evidence to discontinue performing the preventive service

**Study design categories**

- I Randomized controlled trials
- II: Controlled trials without randomization
- **II-2** Cohort or case-control analytic studies
- **II-3** Multiple time series, dramatic uncontrolled experiments
- **III:** Opinions of respected authorities, descriptive epidemiology

General criteria **of effectiveness**

**Screening tests**

Accuracy and reliability of screening tests

Effectiveness of early detection

A, Treatment efficacy

B Asymptomatic period

C Benefits of early detection

D. Acceptability

**Counseling interventions**

Efficacy of behavior change in risk reduction

Efficacy and Effectiveness of counseling patients about health behaviors

Immunizations/ chemoprophylaxis

Efficacy and effectiveness of agent

Efficacy and effectiveness of counseling

SOURCE S H Woolf and H C Sex, "The Expert Panel on Preventive Services Continuing the Work of the USPSTF," *American Journal of Preventive Medicine* 7(5)326-330, 1991

USPSTF, ACP). Generally, priority is given to study methods that are less prone to bias, with evidence from randomized controlled trials rated highest and evidence from sources such as case reports or expert opinion rated lowest (box 7-2) (1 44,207,871).

Different groups use different rating schemes and even among the panels sponsored by a single agency, rating systems may vary. The different guidelines issued by AHCPR, for example, have employed different systems to rate evidence (501 ). Efforts to develop a uniform rating system are complicated by the incorporation of aspects of both the design and quality of the study. Some observers have questioned whether it is always appropriate to give more credence to clinical trials than to other study designs, because a well-done case-control or other quasi-experimental study may sometimes be superior to a poorly conducted randomized clinical trial (376). While explicit rating systems are useful as guides, expert judgment is often still needed to assess the value of many studies.

The USPSTF has sometimes used "causal pathways" to frame the evaluation of evidence (44,717). For example, if evidence is lacking on the association between a preventive service and the outcome of interest (e.g., the impact of screen-

ing adolescents on future scoliosis-related morbidity), the panel examines evidence along the causal pathway (e.g., the relationship between screening and diagnosing scoliosis early, and then the relationship between early intervention and subsequent health outcomes such as back complaints, disability. and psychosocial effects) (figure *7-1 ).*

There are recognized deficiencies in the body of literature available for review. An examination of the literature available on six medical and surgical procedures, for example, revealed:

- few randomized controlled trials on which to rate the procedures' appropriateness,
- incomplete and contradictory information on the indications for and efficacy of the procedures,
- almost no data on costs and utilization, and
- data on complications that failed to specify patients' symptoms or the relationship between complications and reasons for doing the procedure (245).

Consequently, results from studies other than randomized trials often must be used in the guidelines process. Evaluating such studies is another example of the importance of judgments in interpreting evidence. The AHCPR cataracts panel, for example, considered claims data findings that suggested an increased risk of a serious complication (retinal detachment ) in some cataract patients (391). This research lent support to two of the guideline recommendations: that the indications for the procedure be clearly documented in the chart; and that the laser procedure should not be scheduled at the same time as the original cataract procedure (806).[6] Researchers, concerned that limitations of the data used in this study might have led to a misleading finding of complications, are now attempting to collect detailed primary data (724). The AHCPR panel on benign prostatic hyperplasia rejected the use of findings about the risks of transurethral resection of the prostate (TURP) from administrative data, judging that the data source was likely biased ( 140.8 19).

Recognizing that interpretation of evidence may be a matter of judgment, some panels have used formal processes to assess the reliability of these judgments. The AHCPR cataracts panel, for example. had multiple reviewers rate the content and methodology y of research articles and assessed interrater reliability.

## Techniques To Aid Group Interaction and Decisionmaking

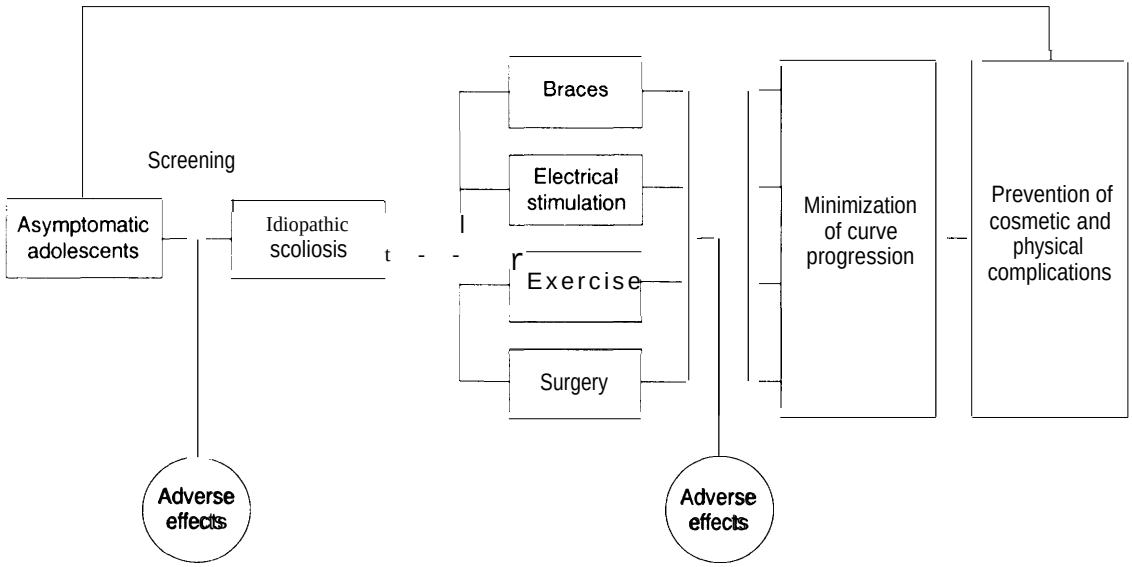
Guideline panels must incorporate information, exchange ideas and opinions, and finally reach some level of agreement on practice recommendations. Group composition and aspects of group process become increasingly important as the availability and strength of evidence declines (469).

How these essential aspects of the guideline development process are accomplished varies considerably. Many guideline processes are informal and have been organized around a series of loosely defined steps:

- A group of appointed experts or representatives is assembled.
- Available literature is collected and summarized (by staff or others) and then reviewed individually by panel members.
- Ideas, opinions. and interpretations of the literature are exchanged in meetings that follow a "roundtable" format. A chairperson facilitates the meetings, often with an appeal to the evidence as it is described in the review.
- Recommendations are made and agreed on, sometimes by a vote (often with a requirement of majority or unanimous agreement).
- Recommendations are reviewed by outside experts and practitioners and then reconsidered by the group.

[6] This last recommendation was partly based on evidence that the laser procedure is not routinely needed, but it was reinforced by the new claims-based information concerning retinal detachment.

## FIGURE 7-1: Examining the Evidence on Screening for Idiopathic Scoliosis Along a Causal Pathway




| Linkage: step in causal pathway | Evidence codes | Quality of evidence |
|---|---|---|
| 1 Accuracy of screening tests. evidence that physical examination of back can detect curves. | II-2 | Fair significant Interrater variation, poor reference standard, lack of evidence form physician screening |
| 2. Adverse effects of screening evidence that screening is associated with an Increased risk of complications, | III | Poor most postulated adverse effects have not been evaluated in studies. |
| 3. Effectiveness of early detection evidence that persons detected through screening have better outcomes than those who are not screened | II-3 | Poor uncontrolled studies based on time trends after initiation of screening, failure to control for confounding temporal factors |
| 4 Braces, | II-2, II-3 | Poor selection bias, lack of internal control groups (most studies), inadequate follow-up, small sample sizes, lack of health outcome measures. |
| 5. Lateral electrical surface stimulation. | II-2, II-3 | |
| 6. Exercise. | I, II-3 | |
| 7, Surgery | II-2, II-3 | |
| 8 Curve progression evidence that curves detected on screening are destined to progress to curves of clinical significance, | II-3 | Fair significant number of patients unavailable for followup, variable measures of progression, |
| 9 Complications of curve progression evidence that persons with scoliosis are more likely to experience back complaints, psychosocial effects, disability. | II-3 | Poor studies generally lack control groups, have high attrition rates, include mixture of patients with different problems, and use variable measures to judge outcome |
| 10 Adverse effects of treatment evidence that treatment is associated with an increased risk of complications, | III | Poor most postulated adverse effects have not been evaluated in studies, |

SOURCE U S Department of Health and Human Resources, Public Health Service, Off Ice of the Assistant Secretary for Health, Off Ice of Disease Prevent[on and Health Promotion, U S Preventwe Services Task Force, Washington, DC, 1993

This approach typifies, for example, the NHLBI and ACIP guideline processes.

### *Formal Group Processes*

A very few organizations issuing guidelines use formal. structured interactive group techniques to orchestrate the guideline process and to make explicit recommendations. Of the groups reviewed here, two-HCHP and RAND, both private guideline developers—use formal group processes. A range of group process techniques have been developed to facilitate group decisionmaking.[7] Some methods are best suited for identifying problems and establishing objectives. Others are designed to help conceptualize alternatives, while still others are tailored to groups that need to make choices among a range of alternatives (538). For the production of any guideline, then, different group processes might be used at different stages of the development process.

Two group processes extensively studied and used to develop clinical practice guidelines are the Nominal Group Technique (NGT) and the Delphi technique (box 7-3) (366). These methods help ensure participation of all members, and they provide explicit decisionmaking rules. Group judgments achieved through either the NGT or Delphi technique generally improve judgments relative to those derived by taking the average of individual judgments, but neither technique clearly seems to outperform the other (660). Both techniques are superior to informally interacting groups in generating new ideas (156).

Formal structured methods can potentially improve group performance by organizing complex information for group consideration, facilitating agreement and decisionmaking, and increasing personal satisfaction of group participants (156). In the absence of a formal process, groups may not perform optimally because one or a few individuals can easily dominate discussions. thereby suppressing the consideration of a balanced set of options. Informal group discussions can also sometimes lack focus and be time-consuming and unproductive. Because much of the expense of guideline development lies in the assembling of experts, methods to make their time together more efficient are desirable.

A potential barrier to using formal methods to structure group process is their unfamiliarity to clinicians. Also, the relative value of these techniques has not been assessed in the context of practice guidelines.

### *Decision Support Systems*

Another way to structure the guidelines process is to use a structured, quantitative framework for integrating and weighting medical and other scientific data. Such support systems can make unwieldy problems more manageable by structuring thought processes, clarifying interrelationships among important factors, and integrating complex data (681). Decision support systems require explicit definitions of the problem, assumptions, events, and outcomes. Such a process helps to assure that relevant factors are considered. and it enables others to review and check the reasoning behind decisions.

Decision support systems help overcome the inherent human limitations of processing information and making judgments (68 1). For example. most people:

- cannot consider more than three to seven alternatives concurrently;
- have a limited cognitive capacity to revise judgments; and
- have biases that affect judgments (e.g., people consistently overestimate the probabilities of events familiar to them and underestimate the probabilities of unfamiliar events).

These limitations in judgment affect the assessment of probabilities, integration of new or contradictory information, estimation of the validity of evidence, and assessment of preferences and values (68 1,942).

[7] Much of the developmental work on group processes has occurred in nonmedical settings (855).

## BOX 7-3: Formal Group Process Techniques Used in Developing Guidelines

**Nominal Group Technique**

The Nominal Group Technique (NGT), developed by Andre Delbecq and Andrew Van de Ven in 1968, has been used widely in human services organizations, business, and as part of evaluative research, The NGT splits problem solving into two phases, an idea-generating phase and a decisionmaking phase A different group process is used for each phase, In the first phase, each member of the group Individually makes a list of ideas for group consideration. This aspect of the process gives the technique its name—individuals participating in the "nominal" group process are a group "in name only" (i e., nominal) and do not initially interact verbally All individually generated ideas are then recorded on a flip chart for the group and are openly discussed In the second decisionmaking phase of the process, individuals vote on priority ideas and a group decision is mathematically derived through rank ordering or rating

**The Delphi Technique**

The Delphi Technique was created at the RAND Corp. in 1950. It was originally used to forecast technological developments, thus, like the Delphic oracle, it was used to look into the future The technique was designed to help groups of experts identify a range of possible program alternatives, explore underlying assumptions or information leading to different judgments, and to reach consensus on complex Issues,

Unlike the NGT, the Delphi technique does not require that participants meet face-to-face, Generally the technique IS typified by the followlng process:

- A questionnaire is dlstributed by mail to a respondent group,
- Respondents independently answer the questionnaire and return it,
- Responses are summarized and a feedback report is developed for each respondent
- Respondents receive the feedback report with a new questionnaire and independently evaluate their earlier responses,

Other factors that affect judgment relate to how questions are framed, For example, clinicians appear to make different judgments in evaluating an individual patient as compared with considering a group of similar patients. Physicians seem to give more weight to the personal concerns of patients when considering them as individuals and more weight to general criteria of effectiveness when considering them as a group (629).[8]

In another example of the importance of framing questions, RAND researchers first asked panelists to rate the appropriateness of certain scenarios for endoscopy and cholecystectomy, specifying that the patients in the scenarios had no comorbidities. When the panelists were asked to rerate the scenarios, with the patients described as having high comorbidity, only of few of the scenarios originally designated as appropriate re-

[8] In one experiment, physicians were asked to make clinical decisions after reading two clinical scenarios. The scenarios were identical except that in one they were asked to evaluate an individual patient and in another they evaluated a comparable group of patients. Physicians were more likely to recommend additional testing and recommend therapy when they evaluated the individual versus the group scenarios (629).

**BOX 7-3 continued: Formal Group Process Techniques Used in Developing Guidelines**

- Respondents are asked to independently vote on priority ideas Included in the second questionnaire and return their responses
- A final summary and feedback report is sent to the respondents and to decisionmakers

The Delphi process often varies according to whether the respondent group is anonymous, whether open-ended or structured questions are used to obtain Information for the respondent group, how many Iterations of questionnaires and feedback reports are needed, and what decision rules are used to aggregate the Judgments of the respondent group The theory underlying the Delphi technique is that Improvements in judgment with each Delphi Iteration occur because the most knowledgeable panelists confidently retain their Judgments and anchor the median close to the true value, while less knowledgeable panelists change their Judgments to be closer to the median If this in fact occurs, the median response should move toward the truth over rounds of the Delphi process

Those who maintain their same judgments over Iterations may not in fact be more knowledgeable, but instead be dogmatic and intransigent If so, the convergence of opinion observed over Iterations may just reflect the influence of a dominant individual To alleviate this effect, substantive feedback to panelists must include not just median ratings but also justifications for ratings based on the evidence at hand There is a tendency for group judgment to converge over time and without appropriate feedback, such convergence could represent an artifact of the method rather than true convergence of opinion

SOURCES Office of Technology Assessment 1994 based on information from A L Delbecq A H Van de Ven and D G Gustafson *Group* Techniques *for Program Planning A Guide to Nominal Group and Delphi Processes* (Glenview IL Scott Foresman and Company 1975) G Rowe G Wright and F Bolger Delphi A Reevaluation of Research and Theory *Technological Forecasting and Social Change* 39235-251 1991

mained so designated ( 13 percent for endoscopy, 33 percent for cholecystectomy) (405).

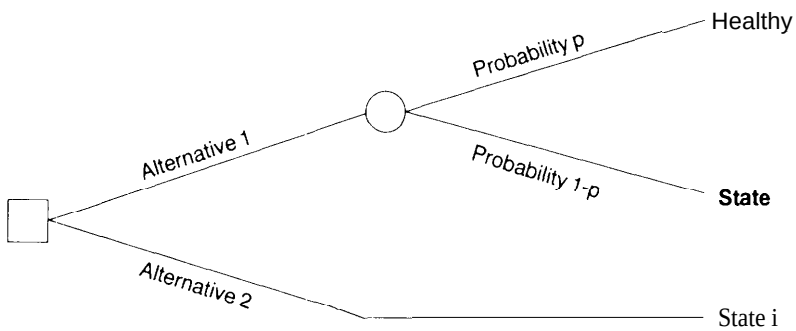
One of the most common decision support systems used in developing guidelines is *decision analysis,* a useful structure for determining the preferred course of action under conditions of uncertainty *(547),* Decision analysis provides a framework for specifying the probability that a particular clinical state exists and quantifying the value of the various outcomes of a decision (see chapter 3 for a more detailed description of this technique). A decision analysis attempts to answer the questions. “Is it more desirable that I do this or that’?” and “If this is so, what is the probability that that is so’?” It is used during the guideline panel’s attempt to consider all the relevant information. Steps involved in decision analysis include defining all possible outcomes of interest. quantifying their probability of occurring, and sometimes considering the costs and benefits associated with each outcome (box 7-4).

(Algorithms are a related framework that are sometimes used to assist the clinical decision maker actually using a guideline. In contrast to decision analysis, algorithms prescribe, "Given this. do that” (490) (see box 7-1 ).)

Although they are theoretically attractive and can be very useful, decision models also have limits. Generally, the systems are complex and time consuming. Specification and structuring of the problem, obtaining the values for the data inputs, and computation of the primary and subsidiary analyses require substantial expertise in clinical

## BOX 7-4: Steps in Using Decision Support Systems

- Identify the problem in terms of the clinical presentation, population, time frame, and perspective (e g patient, payer, provider, society) The perspective of the model affects the costs and values assigned to events and outcomes and thus strongly influences the results.
- Structure the problems and explicitly describe the underlying logic and reasoning. Alternate courses of actions and their consequences must be specified. Outcomes of interest might include physiological parameters, such as mortality/survival and complications, and physical, social, and psychological, cognitive, role, social, and other functional measures Broader measures that incorporate patient values and preferences and costs, such as quality adjusted life years, cost-effectiveness, and cost-benefit also should be considered. Based on the outlined problem structure, the probability of occurrence of events and outcomes must be obtained from objective, published, peer-reviewed scientific literature, but may be based on expert judgment when other data are unavailable. The model is often expressed as a decision tree with branches representing different outcomes




- Select preferred options considering the expected value of each alternative strategy conduct sensitivity analysis when data underlying the decision model are uncertain to assess the likely range of values associated with options.

SOURCE J S Schwartz "Decision Support Systems and Their Potential Contribution to Consensus Development paper presented at "Workshop To Improve Group Judgment for Medical Practice and Technology Assessment, " sponsored by the Institute of Medicine, Division of Health Care Services, Council on Health Care Technology, Washington, DC, May 15-16, 1990

medicine, epidemiology, biostatistics, economics, psychology, and decision sciences. Many times, data required to model decisions are not available. Also, models only inform decisions; they are not definitive.

There have not been many scientific evaluations of the impact of using decision analysis in group judgments. The limited experience that does exist suggests that for decision support systerns to work, the group must be receptive to the concept. The technique is unfamiliar to many, so it must be taught to potential users (681).

### *Decisionmaking Rules and Procedures*

A number of methods to combine the opinions of individuals in a decisionmaking group are avail-

able, some of them employing sophisticated mathematics and weighting schemes. These different methods may give widely different answers for certain questions (942).

Some guideline panels require the consensus of group members. while others allow for a range of dissenting opinion. Consensus does not necessarily mean unanimous agreement. In fact, it can be taken to mean group solidarity in sentiment and belief. a general agreement, or the judgment arrived at by most of those concerned **(855,899).** Most of the guideline processes reviewed for this report use informal consensus methods to arrive at recommendations. Groups consider evidence and usually iron out differences in roundtable discussions, but sometimes vote when there is disagreement. Few groups require unanimous approval of the guideline, and somc have established mechanisms to include dissenting opinions into guideline reports (e.g., ACP, HCHP).

Some observers suggest that requiring unanimity may result in recommendations that represent the 'lowest common denominator" of opinion. Instead, levels of agreement or disagreement can be establishcd according to votes taken during the group process. This provides a mechanism to voice disagreement without endangering the overall group process ( 123,51 ()). Voting can be either anonymous or public. If anonymous, those who are in the minority have some protection from undue pressure to change their position. Public votes may allow the group to focus on the problems that remain to be resolved (or that cannot be resolved) and force dissenters to defend their positions. Voting can be done On a simple yes/no basis or on a scale that reflects the level of agreement or disagreement. Using a scale allows panelists more latitude in expressing their opinion and can be used to qualify any recommendations according to strength of opinion (5 10). Dissenting opinions can also be included in the final report ( 123).

The RAND process allows participants to rate appropriateness indications anonymously. Some fear that the RAND process may lead to conclusions that diverge from the medical literature because of the nature of group process. For example, the reduction in disagreement over Delphi iterations could be the result of well-known psychological pressures toward conformity in groups, or a methodological artifact resulting from statistical regression to the mean (949).

However, some evidence suggests that conflict resolution in groups is determined more by the availability of research evidence than by the personalities and predilections of panel members (469). There arc limits to the extent to which agreement can be reached when good evidence is lacking. Nearly three-quarters of conflicts were resolved during a consensus process when good data were available, while only about one-quarter of conflicts were resolved when good data were not available (469).

The outcome of group processes can be enhanced if sources of disagreement are identified and discussed (5 10). It is informative, for example, to know whether disagreement stems from some panelists' concerns about a poor health outcome or from perceived unfavorable patient attitudes toward the intervention under consideration.

Not surprising] y, how agreement is defined can greatly affect a panel level of agreement. RAND assessments of appropriateness for coronary angiography, for example. ranged from 31 to 63 percent depending on how agreement was defined, and whether some panelists opinions that represent extremes were discarded in the final judgment (586).[9,10]

[9] The purpose of discarding opinion outliers is to better represent the group view. It is not necessarily the case, however, that the outlier is wrong (684).

[10] There was agreement for 31 percent of coronary angiography cases when all nine of the ratings fell within any three-point range. There was agreement for 63 percent of the cases when after discarding one extreme high and one extreme low rating, the remaining seven ratings all fell within any three-point range (586).

## The Basis for Guideline Recommendations

Professional judgment is used throughout the guideline development process—from reviewing and interpreting key evidence to discussing personal opinions and experience and formulating recommendations. There is, however, great variation in the extent to which expert opinion or judgment is used as the basis of guideline recommendations. At the extreme is NCI, which has recently decided against issuing recommendations at all. Panels apply their judgment in evaluating and summarizing the available literature, but conclusions are limited to scientific statements that do not explicitly promote particular clinical policies.

The prevention guidelines issued by the U.S. Preventive Services Task Force in 1989 set a benchmark in the use of evidence to support guideline recommendations (871 ). Unlike any previous U.S. guideline efforts, the task force prefaced their work with the development of an explicit approach to selecting and evaluating the existing literature, as described above (see box 7-2). Further, their guidelines graded the strength of each recommendation according to the strength of the evidence supporting it. The USPSTF was the first major officially sanctioned group to produce practice guidelines linked directly to evidence, and its efforts were fundamental in establishing their practicality and acceptance. ACP's CEAP program also produces evidence-based guidelines.

Like the Preventive Services Task Force guidelines, AHCPR's guidelines effort was established with the intent of applying an explicit, systematic approach to the selection and evaluation of evidence regarding the effectiveness of managing a spectrum of health conditions (812). Unlike the USPSTF effort, however, AHCPR has a separate guideline panel for each condition selected for assessment, and different panels have interpreted and carried out this task in different ways (678). Where there has not been a strong evidence base on a topic, panels have sometimes made recommendations primarily on the basis of clinical opinion. The panel for the AHCPR guideline on pressure ulcers conducted a systematic review of the literature, found few quality studies, and so based guideline recommendations on the expert opinion of the panel.

The NIH Consensus Development Conference Statements are based on the consensus of opinion of a group of nationally recognized experts who consider evidence presented to them over the course of a three-day meeting. The panelists review clinical evidence and use their best judgment to make recommendations. There is no attempt to link recommendations explicitly to a particular source (either the oral presentations or published literature). Similarly, physicians rating the appropriateness of interventions at RAND use a combination of scientific evidence and expert judgment. NHLBI's cholesterol guidelines rely heavily on indirect laboratory evidence and expert opinion in addition to considering evidence from clinical studies.

Those supporting a role for expert opinion in forming guideline recommendations argue that where good evidence does not exist, the best judgment of experts is at least better than no guidance for clinicians at all. Advocates of this position also point out that there are many circumstances where decisions have to be made in the absence of good evidence. For example, HCFA has had to make many decisions regarding coverage of specific technologies under Medicare on the basis of expert opinion alone (188). In contrast, those supporting a strong evidence-based approach argue that where good information is lacking, there is no sound basis for creating or promoting a position. Without the benefit of strong scientific evidence, groups of experts can come up with very different recommendations, making the guidelines unreliable, One researcher, for example, identified 21 guidelines on asthma management which varied greatly in content (277).11 Indeed, guide-

[11] The guidelines were identified through the National Library of Medicine's MEDLINE® database covering the period 1989 to mid-1993.

lines can differ even when some evidence exists. Past experience with guidelines+. g., guidelines on screening for colorectal cancer (box 7-5)-confirms that conflicting recommendations frequently occur.

For some interventions, it may not be possible to produce good evidence quickly. For many preventive interventions for chronic disease, for example, the length of time between intervention and potential outcome is so long as to make randomized controlled trial results unavailable to policymakers for many years. Studies of how best to diagnose and treat some acute conditions, however, can often be conducted relatively quick] y. In a recent example. decision rules for the use of radiography in acute ankle injuries were developed and then assessed in a controlled trial. Investigators were able to link implementation of the decision rules to a decrease in use of ankle radiography, waiting times, and costs without patient dissatisfaction or missed fractures (502,723).

Even when good evidence from randomized controlled trials is available, clinician judgment is needed to interpret the results. For example, often trials may be confined to a limited group of patients (e.g., middle-aged males) and clinicians need to assess whether the trial results could safely be generalized to other groups of patients (elderly people, women, etc. ) (378). And just because relevant evidence from clinical trials is available does not mean that the interpretation of results is always clear cut. There are many instances of contradictory results from multiple randomized clinical trials on the same topic. Sometimes these differences reflect deficiencies in design (e.g., they were too small), but in other cases conclusions from well-designed trials on the same topic differ because of variations in patients or therapies across trials (e.g., trial entry criteria or treatment of placebo groups may differ somewhat across trials) (355). Thus, even when several trials are available. there is often room for variations in opinion on the interpretation of evidence. For example, vigorous debate over whether mammography is indicated for women under age 50 followed the recent pub] i cation of results of a large randomized controlled trial designed to assess the effectiveness of mammography in reducing breast cancer mortality (37, 109,255,696).

Notwithstanding the fact that interpreting even good evidence is itself a matter of judgment, studies of group processes show that agreement among panel members is easier to reach when good evidence is available (469). Panel ratings of appropriateness of indications for Cesarean birth were much more likely to be in agreement for indications for which there was good evidence (e.g., randomized controlled trials or other prospective studies) than for those for which evidence was lacking or of poor quality (469). Likewise. groups are more able to make clear and precise recommendations when good evidence is available and agreed upon (469).

There are limits to the extent to which agreement can be reached when good evidence is not available. For example. when good evidence on the appropriateness of cardiovascular procedures was available, physicians in the United Kingdom and in the United States generally agreed, but the U.K. panel produced consistently lower ratings for many indications when evidence was unavailable. The U.K. panelists also seemed to require a higher standard of scientific evidence than did their U.S. counterparts (85).

How clinicians draw conclusions about appropriateness in the absence of evidence is very unclear, In one study, physicians rating the appropriateness of endarterectomy appeared to base their judgments mom on a patient's risk status prior to surgery than on their assessment of how the procedure would change outcome (i.e., probability of death or stroke). For example, even though a panel of expert clinicians assessed six indications for endarterectomy as inappropriate, six of eight panelists believed the procedure reduced the likelihood of adverse outcomes for these indications (500). This seeming contradiction is perhaps explained by the apparent importance of surgical risk in rating appropriateness. Patients at high surgical risk were often assessed as inappropriate candidates for the procedure in question.

### BOX 7-5: Conflicts and Issues in Federally Sponsored Guidelines for Colorectal Cancer Screening

Several federal agencies sponsor the development of clinical practice guidelines for preventive services, and two have produced guidelines relating to colorectal cancer (CRC) screening The recommendations in these two guidelines differ from each other and serve as an illustration of the potential for contradiction among multiple guidelines. They also differ from the privately sponsored guidelines issued by philanthropic groups and by physician-specialty societies The various groups Issuing guidelines on this topic, and some of the differences among them, include

- the U S Preventive Services Task Force (USPSTF), which has declined to recommend either for or against periodic screening with either fecal occult blood testing (FOBT) or sigmoidoscopy in average-risk individuals,
- the National Cancer Institute (NCI), which has recommended an annual FOBT and a sigmoidoscopy every three to five years starting at age 50, with no suggested age at which to discontinue screening,[1]
- The American College of Physicians, whose recommendation is similar to that of NCI,
- the American Cancer Society which likewise recommends frequent FOBT and sigmoidoscopy but in addition recommends an annual digital rectal exam after age 40 and two initial sigmoidoscopies one year apart at age 50, and
- the American Society for Gastrointestinal Endoscopy and the American Gastroenterological Association, which endorse both FOBT and sigmoidoscopy screening beginning at age 50, but have not provided a recommended frequency.

The differences among groups in recommendations regarding CRC screening for average-risk people reflect two facts First, the evidence on the effectiveness of specific technologies is inadequate in many areas Second, the criteria (either implicit or explicit) for judging the evidence that does exist differ among expert groups.

At issue is whether a screening test for CRC must be shown to reduce CRC incidence or mortality in order to be considered effective, or whether demonstrating a shift in the distribution of

[1] NC I recently ceased issuing or endorsing recommendations for cancer screening and limits Its statements to reviews and interpretations of the evidence (see chapter text and appendix C)

even though clinicians, when asked to assign outcome probabilities to such patients, indicate that they would likely improve with surgery (500).

## Organization and Administration of Guideline Activities

Almost all federal guideline activities described here are sponsored by the government rather than being developed internally. This distancing from the sponsoring agency provides a measure of independence. The guidelines of AHCPR and NIH, for example, do not need to be formally approved by the sponsoring agency. [12] The sponsoring agencies do, however, play important roles throughout the guidelines process, for example, in selecting topics and panelists, specifying methodology, and providing administrative and technical support to the panel. CDC guidelines published in *Morbidity*

[12] AHCPR reserves the right to publish AHCPR-sponsored guidelines and seeks both agency and departmental clearance (53).

**BOX 7-5 continued: Conflicts and Issues in Federally Sponsored Guidelines for Colorectal Cancer Screening**

detected cancers to earlier stages IS sufficient for considering a screening regimen effective Those who require direct evidence that CRC screening wiII reduce the Incidence of, or mortality from, CRC have found the existing evidence inadequate, The critics also point out that screening and diagnostic followup have medical risks and high costs Others focus on the heavy burden of illness and death brought about by CRC and conclude that even indirect evidence that screening may alter the course of disease in a substantial proportion of people screened cannot be ignored

The controversy around guidelines for CRC screening is likely to continue for some time The USPSTF IS currently updating its recommendations, including those for CRC screening At the same time, the Agency for Health Care Policy and Research (AHCPR) has recently begun sponsoring the development of its own guidelines on the topic The agency awarded the contract to develop the guideline to the American Gastroenterological Association (AGA) There is no formal mechanism for coordinating the USPSTF and AHCPR efforts, and history suggests that there IS considerable potential for conflicting recommendations between the two forthcoming sets of recommendations

The potential for conflicting recommendations is heightened by the fact that the panels creating the new CRC guidelines are likely to be quite different and to operate in different ways The USPSTF includes no gastroenterological specialists on its panel, and as described in the text it follows a rigorously structured process of considering evidence and developing recommendations In contrast, the AHCPR-sponsored panel is not required to follow any equivalent development process under its contract The contract does specify some of the procedures the AGA must follow in appointing panelists (e g , panel members must represent consumers as well as a variety of health care professionals, and AHCPR wiII review the proposed panelists for "potential conflicts of interest, but the contractor has considerable leeway in deciding exactly who will be on the panel Panel composition and the contract award may be an issue in future debates about the panel's recommendations, particularly if those recommendations differ from the recommendations of other publicly and privately sponsored groups

SOURCES Office of Technology Assessment *Costs and Effectiveness of Colorectal Cancer Screening in the Elderly* (Washington DC U S Government Printing Off Ice September 1990) U S Department of Health and Human Services Public Health Service Agency for Health Care Policy and Research contract no 282-94-2023 awarded to the American Gastroenterological Association May 25 1994

*and Mortality Weekly Report, however,* are ap - proved by CDC and represent government policy (751).

Some guideline panels are, in effect, standing committees that assume a long-term commitment to a particular topic. The USPSTF, NCI, and the CDC's ACIP panels follow this format. NHLBI activities are overseen by standing "Coordinating Committees." Each committee is charged with staying abreast of scientific developments and monitoring health education needs in its particular area. The committee can initiate a variety of activities from creating health education brochures to establishing panels to develop clinical guidelines. Standing committees have the advantagc of being able to keep abreast of the 1iterature on a given topic after a guideline has been published to decide when the guideline needs to be updated. The NCI has a formal process to continually monitor and update the information statements on the PDQ

computerized database. This medium has advantages over guidelines that may become out-of-date soon after they are published.

## Methodological Research

Several groups have cited a need for further rew arch on the processes that underlie guideline development (37 1,376,607), but relatively few studies have judged the quality of guideline development processes. of the studies available, none compare the relative merits of one method to another.

Instead. research to date has focused on individual approaches—in particular, the RAND appropriateness method. As discussed above, research from this source on panel methods has illuminated the importance of such characteristics as panel membership. de fin it ion of agreement, and availability of evidence on the reliability of guideline results. RAND researchers have also found individual physicians able to be consistent across time in their recommendations. Physicians who rated appropriateness were able to reliability reproduce their ratings six to eight months after the completion of the RAND process (526).

In another interesting experiment, performed at HCHP, three panels of primary care internists were provided with identical literature summaries on the management of two common clinical problems: acute sinusitis and dyspepsia. Each panel used formal group processes to create clinical algorithms. Five of the six algorithms produced by the panels were similar, but one was substantially different. The authors concluded that "even with optima] literature support and a standardized consensus process, physician consensus groups may still produce guidelines that vary due to differences in interpretation of evidence and physician experience." Evidence available to the panels included a wide range of studies, with varying degrees of epidemiologic rigor and some conflicting results. The topic about which the panels' disagreement was greatest was not addressed in the available literature (594).

Guideline development methods used by the federal government have not been formally assessed for reliability y or validit y. The NIH Consensus Development Conferences have been evaluated only for their effects on physician practices. CDC has created a database of CDC-developed guidelines and developed resource materials relating to decision and cost-effectiveness analyses but has not evaluated the methods of developing guidelines themselves.

AHCPR has begun assessing its process of developing guidelines, and it is sponsoring a study on optimal methods for prioritizing guideline topics (813). Investigators at RAND are evaluating differences between their appropriateness rating method and the guideline methods of AHCPR (689). In addition. RAND researchers are investigating (403 ,823):

- the use of meta-analysis in the literature review,
- the effect on appropriateness judgment of having the panel consider probabilities and utilities explicit] y,
- the effects of alternativc methods of panel composition and function,
- methods to evaluate service underuse,
- the reliability y and validity of panel ratings, and
- the relationship between patient outcomes and inappropriateness ratings (81 ,373).

# CONCLUSIONS

## The Link Between Methods and Recommendations

The methods used to develop clinical practice guidelines might be a relatively uninteresting topic if the guidelines issued by the various organizations were consistent and uniformly accepted as valid. As previous examples demonstrate, however. they are not. Recommendations by various groups conflict with each other, and the recommendations of one group on a topic often are not considered valid or acceptable by others. Differences among guidelines can cause confusion and may undermine the basic credibility of guidelines themselves (205).

Although the focus of various federal guideline efforts vary somewhat from each other, there is no overall federal guideline strategy or coordination,

and current efforts are fragmented. For example, AHCPR has issued guidelines on topics also covered in the National Institutes of Health's Consensus Development Program (e.g., urinary incontinence, pain management). The U.S. Preventivc Services Task Force, sponsored by the Office of Disease Prevention and Health Promotion. and the Centers for Disease Control and Prevention have both issued guidelines on immunization. Recommendations regarding cholesterol screening are issued by both the USPSTF and the National Heart, Lung, and Blood Institute. Government-sponsored efforts sometimes also overlap with private sector activities. The American Academy of Pediatrics, for example, also issues immunization recommendations, as does the ACP.

The diversity of organizations producing guidelines and the methods they use-even within single agencies, such as NIH—suggests that **the potential for unnecessary duplication and contradiction between guidelines, and inefficient cross-agency use of resources needed to produce guidelines, is high. Furthermore, recommendations of guidelines available on the same topic sometimes differ markedly.**

Guideline methods vary considerably across federal agencies, yet there have been few efforts to compare them and identify the relative strengths of competing approaches.

The limited research on guideline development processes generally suggests that the availability of strong, high-quality evidence improves the likelihood that panels of experts will agree on practice recommendations. Group composition and aspects of group process become increasingly important determinants of guideline recommendations as the availability of evidence declines.

How evidence is considered, how group discussions are managed, and how agreement is defined also appear to affect the decisions that groups make. Consistency of methods appears to improve the reproducibility of guidelines, but where evidence is lacking the differing judgments of panel members place limits on the ability to produce similar guidelines even when similar methods are used. 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. More research is needed to identify the factors that affect group judgments when evidence is lacking.

The composition of a guideline panel appears to affect the scope of a guideline, the kinds of issues considered, the way in which panel members consider and weigh different types of evidence. guideline recommendations. and the credibility of the guideline. There are probably tradeoffs in the effects of panel composition regarding such issues as credibility by physicians vs. considering topics such as cost and patient preferences, but the implications of these tradeoffs have not been adequately explored.

Evidence-based clinical practice guidelines have proved workable and politically acceptable, **In fact, the theoretical strength of such guidelines is so compelling that it calls into question the usefulness of federally sponsored guidelines not based on an explicit review of evidence that considers, in some explicit and systematic fashion, the strength of that evidence.** Guidelines with less evidence basis may be justified for some purposes (e.g., guidance on the use of very new technologies), but those purposes should be carefully thought out.

The identification of outstanding clinically relevant questions for research is an important contribution of guidelines, and such recommendations could be highlighted to a greater extent.

## Developing Methods To Prioritize Guideline Topics

Guidelines are most likely to be influential when sound evidence supports certain clinical practices, but clinicians are not following those practices because they lack information or are uncertain. **Priorities for guidelines could be set according to formal reviews of available guideline-related evidence** *and* **analyses of clinical practice.** If guideline topics are selected solely on the basis of practice variation (or some other indicator of prac -

titioner uncertainty), there will likely be many topics selected that have an insufficient pool of evidence with which to develop a clear and specific guideline. It is difficult for groups to formulate specific guidelines useful to practitioners when evidence is not available and without good evidence, panel judgments seems to be vulnerable to significant bias.

Most groups developing guidelines select a topic and then proceed to assess the available evidence. Criteria could be established to help determine when sufficient data are available to develop a guideline (377). Such criteria are generally not being used now. Some suggest that the 1984 NIH recommendations on lowering blood cholesterol may have been issued too early, before adequate information was available on which to make recommendations. Evaluation of the state of evidence prior to guideline development will be easier with the establishment of a NIH clinical trial database (see chapter 4).

**If the intent of a guideline is to inform and possibly change physician behavior, data could be collected prior to guideline development to evaluate the state of medical practice, and clinician knowledge, attitudes, and beliefs.** Such data could be used to assess whether guidance is needed, and if so, if it is likely to be adhered to by the targeted population. Had the state of current practice been assessed, an NIH Consensus Development Conference on treatment of primary breast cancer might not have been held. A major recommendation of the conference—that few Halsted radical mastectomies be done—was found in a subsequent review of medical practice to be moot-the procedure was being performed very infrequently (372,41 1).

Behavioral science techniques such as focus groups and surveys might be helpful in understanding the sources of variation in practice. Both focus groups and surveys have been conducted by the NIH Consensus Development Program, but they are used to evaluate the impact of their program and not to identify topics. The AMA DATTA process assesses expert physicians' opinions regarding the safety and effectiveness of technologies, which could provide useful information to guideline developers. The AMA survey of physicians regarding the use of Teflon™ injections to treat incontinence would have informed the AHCPR guideline on urinary incontinence (the AMA survey was conducted after the AHCPR guideline was published). Most physicians polled in 1992 by the AMA felt that Teflon injections were effective in certain circumstances, but most questioned the safety of the technique (407). (The FDA has approved the use of Teflon preparations for injection into vocal cords to treat paralytic dysphonia. The use of this paste for the treatment of urinary incontinence is considered to be an "off-label" indication.) The 1992 AHCPR guidelines assess the effectiveness of the procedure (it is listed as a treatment option) but not its safety, which seems to be of concern to clinicians (531).

## Research Needs

It is important to establish which processes produce the most valid. reliable, 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 particularly appropriate to certain purposes or under certain circumstances, but at present there is little evidence on which to tailor guideline efforts.

**Some federal guideline efforts have no published description of their process, making it difficult to judge the basis and soundness of the guideline or to compare different approaches. The methods used to make guideline recommendations need to be explicit if any evaluation comparing them is to be made. Aspects of the process that need to be fully described within the** guideline include how the topic was selected, how relevant evidence was identified and considered, how panel members were identified and selected, and how evidence and expert opinion were used in making recommendations.

A number of methods are used to make clinical practice recommendations. but there is insuffi-

cient research with which to judge what method works best (246). Guideline development panels arc now often quite small (generally 10 to 20 members), but take on variety of tasks. They need to conduct literature reviews, organizc and synthesis evidence, evaluate safety and effectiveness of alternative interventions. consider cost-effectiveness and othcr health policy considerations, and identify areas of needed research within the guideline topic. In some ways current guideline panels are inefficient. Given the range of responsi - bilities of a panel, it is difficult to provide the needed expertise on panels of limited size. The process is sometimes very slow and experience gained is often lost after the panel completes the guideline.

One alternative model to test would be to create standing expert teams to support guideline panels. Such teams could experiment and develop methods for activities that all guideline panels must do: literature review; assess current practices: consider the health policy implications of the guideline such as the cost impact of guidelines and the cost-effectiveness of alternative clinical strategies.

**Existing group processes should be further developed and tested and contrasted with one another. Formal group processes seem to enhance group performance, and some cvidencc suggests that guideline panelists prefer structured approaches when the topic is controversial (689). In** the area of guidelines, the RAND adaptation of the Delphi group process method is a tested applicable formal group process model. This model structures group interaction and formalizes the elicitation and combination of' member opinion, but the foundation of that opinion is not specified. There is no way to know the extent to which assessments are based on opinion or evidence. The RAND method includes a review of relevant literature, but appropriateness ratings are not linked to the evidence. One way to strengthen their approach would be to identify the source of each rater opinion. This would be difficult at present, given the large number of indications that need to be rated. How'ever. most indications that are identidied are theoretical and are rarely or never seen in practice, [13] One could therefore consider asking panels to provide a richer set of judgments on fewer indications.

Federal agencies are in a unique position to be able to assemble resources needed for guideline development. In addition to funding clinically relevant research to serve as the basis for guidelines, agencies could develop tools potentially useful to guideline developers. These might include:

- identification of areas of clinical uncertainty and its sources—national databases can be used to identify practice variation, and national clinician surveys could be conducted to assess sources of variation:
- methodologic reviews of the literature on topics of likely interest—teams of methodologists could review literature identified through systematic searches;
- development of methods to incorporate cost assessments and patient preferences into practice guidelines; and
- listings and evaluations of clinical trials.

The planned creation of two clinical trial databases. if successful, will facilitate the identification of evidence for guidelines panels. All published randomized controlled trials will be identifiable through a database being developed through the efforts of the National Library of Medicine and the Cochrane Collaboration (see chapter 4). In a separate effort, NIH is creating a database of NIH-sponsored ongoing trials related to women's health, and the possibility of an all-NIH database is also under consideration (chapter 4).

[13] As many as 60 percent of indications for angiography, 79 percent of indications for endoscopy, and 68 percent of indications for endarterectomy were never seen in reviews of patients' charts (586).

# The Impact of Clinical Guidelines on Practice 8

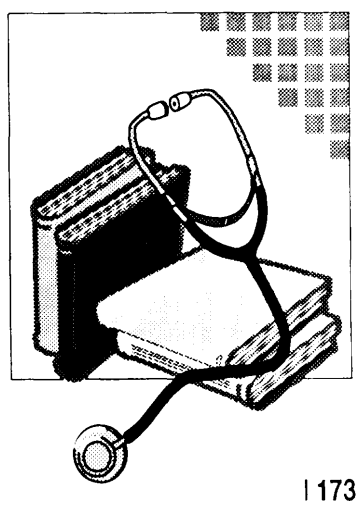
To improve health outcomes and health care costs, the evidence produced from evaluations of health technologies and practices must be disseminated to providers and patients in a manner that will convince them to adopt new findings and recommendations into everyday decisionmaking. This chapter focuses on the impact of evidence. and specifically evidence assimilated and presented as guidelines, on clinical practice.

The chapter first outlines alternative behavioral models of clinical decisionmaking, the major forces that shape physician practices, and some of the strategies that have been used to change behavior. ' It then examines the interventions used to implement clinical practice guidelines and discusses factors that can affect their impact.

## ANATOMY OF PHYSICIAN BEHAVIOR

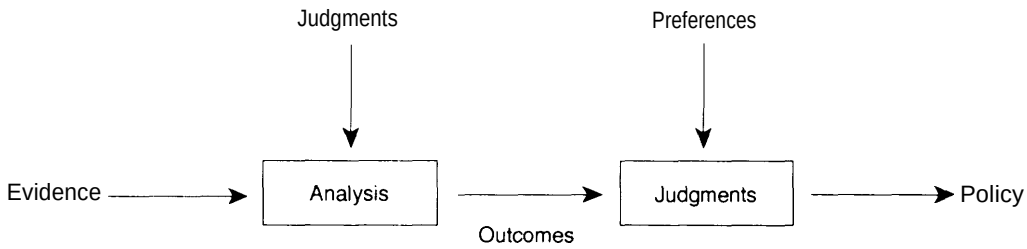
### Modeling Clinical Decisionmaking

Behavioral models of clinical decisionmaking can provide insights into the forces that shape individual clinical decisions. Traditional models of clinical decisionmaking have emphasized the role of information as an influence on practice. One of the simplest models of clinical decisionmaking presents it as a two-step process (see figure **8-1 ) (198). In the first step, the decisionmakers consider evidence about the possible outcomes—the benefits,**



[1]This chapter focuses on the physician as the professional clinical decisionmaker, since the available 1 iterature is best developed for physicians. However, many of the principles discussed also apply to nurses, physicians assistants, and other clinical service providers.

**FIGURE 8-1: A Simple Model of a Clinical Decision**




SOURCE D M Eddy, ' ClinicalDecision Making from Theory to Practice—Anatomy of a Decision,' *Journal of the American Medical Association* 263(3) 441-443, Jan 19, 1990 Reprinted with permission

risks, and costs—associated with each possible alternative. If the decision was about the use of mammography to screen for breast cancer, for example, this step would involve considering information on factors such as the relative risk of death from delayed vs. early discovery of breast cancer, the likelihood of a positive finding, the risks of radiation exposure from the test, and the risks of extra biopsies that might be performed if the mammogram is suspicious. In the second step, patients and physicians make judgments about the relative desirability of those outcomes and decide whether they want the screening,

According to this model, the link between evidence and practice is improved by increasing the quality and quantity of evidence available, devoting more effort to soliciting and incorporating patient preferences, and ensuring that physicians and patients follow the decision steps. This perspective underlies the expectation that increasing the supply of results from effectiveness research to physicians will improve the quality of clinical decisions and patient care.

Many simple clinical decisions (e.g., prescribing antibiotics for certain types of infections) can be made according to such a model of physician behavior. The model applies well when the evidence for benefit is clear, patients have no complicating conditions, and preferences for outcomes are predictable.

Many other decisions, however, are more complex and challenging. Complexity arises when good evidence from clinical studies is not available to provide reliable estimates of risks and benefits for a particular type of patient, and when the decision involves the balancing of multiple factors and less predictable outcomes. Challenges can also arise in assessing patient preferences, or in deciding what level of risk for bad outcomes the physician and patient are willing to endure.

Evidence and preferences are fundamental elements of every clinical decision. However, many other forces are also known to affect clinical deci - sionmaking, and more elaborate models are required to explain many individual decisions and broader patterns of practice. These other factors help explain why actual clinical practice often differs from expected or ideal practice, and why evidence and patient preferences alone do not always determine clinical choices.

The more complex models of physician behavior owe much to the work conducted over the past several decades in such areas as adult learning theory, technology diffusion, and communication. Drawing on the studies in these areas, several authors have developed models that view clinical decisions as emerging from a balance of a range of forces in the health care environment (214,240, 466,682). One such model, for example, characterizes physician decisionmaking as influenced

by educational, administrative. economic. patient. community, and personal forces (see figure 8-2) (466).

These more complex models have several important implications for the design of strategies to incorporate new evidence into clinical practice:

1. Evidence is only one of many inputs to decisionmaking and may be overwhelmed by other forces.
2. Simple dissemination of information to physicians may be ineffective in changing behavior.
3. Information may flow to physicians through a variety of channels, and its impact may be increased if multiple information channels are used.
4. In different contexts. one or another force may be dominant, suggesting that strategies for changing practice need to be tailored to fit the circumstances of the physician, the medical condition, or the type of practice involved.
5. Since the forces that influence practice are always active, the ability of interventions to produce lasting change is the best measure of their value (466).

## Sources of Influence on Physician Practice

In order to design interventions that are efficient and effective for communicating new information and changing behavior. it is useful to develop a clear understanding of the full range of factors that underlie current practices. The factors that may affect physicians" willingness to change their practices should also be considered.

**Prevailing theory. The conceptual foundation underlying current** practice serves as the basis for incorporating new information, and it will affect the readiness with which new information is received (240).

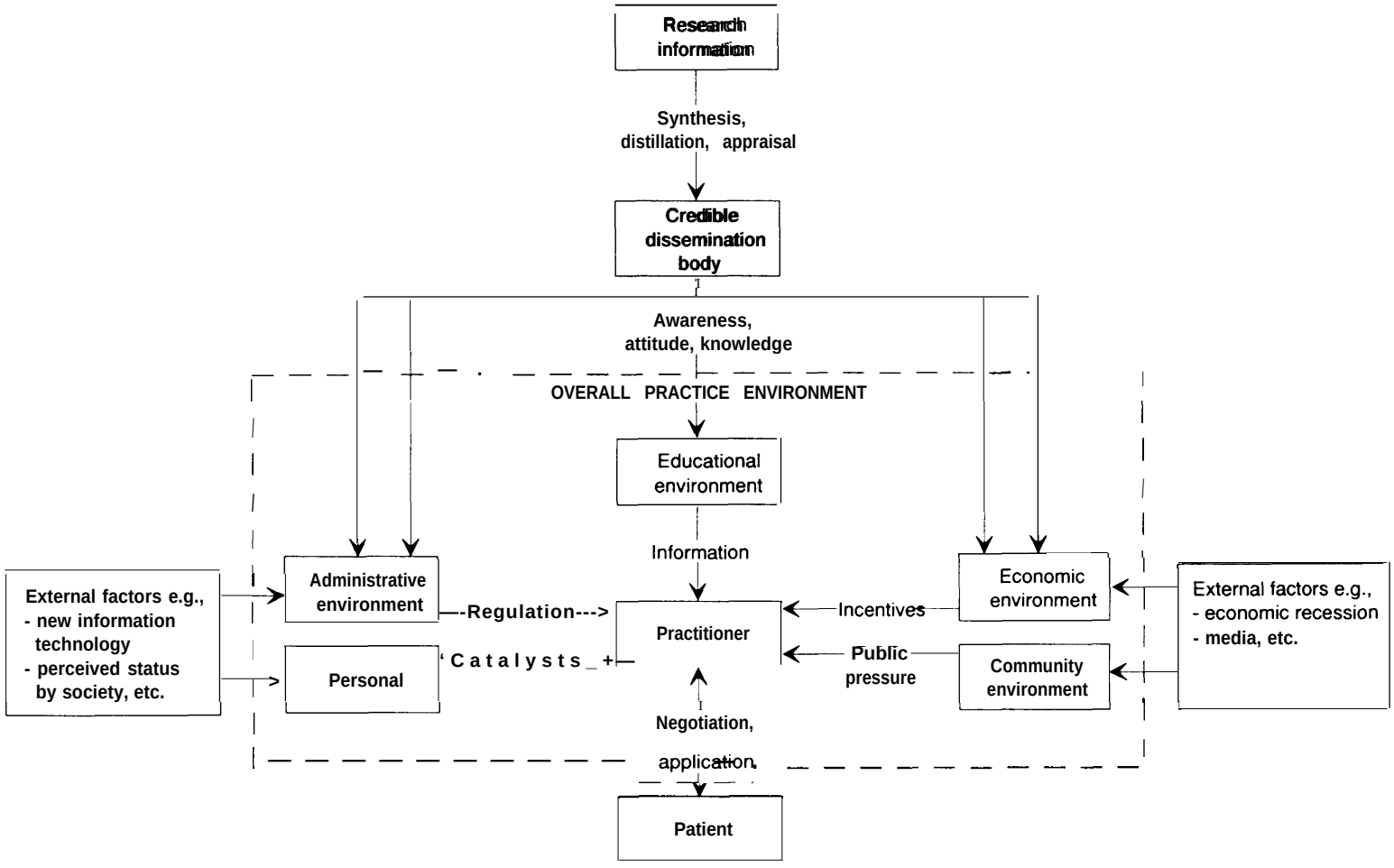
Nature of the innovation. The nature of the change in practice influences the extent and rapidity of change (240). Physicians are much less likely to adopt new practices that involve major changes from existing practice than those that involve only slight modifications (265). New behaviors are also more likely to be adopted if they are simple and relatively inexpensive. represent major and easily observable improvements over current practice, are easily understood according to accepted principles, build on existing skills and abilities. and are usable on a trial basis and easily abandoned if found inadequate (644.73 1 ). Physicians may also be slow to incorporate recommended practice changes if they do not have the necessary skills or resources to implement them.

**Type of clinical decision.** The ability to change existing practice depends not only on the characteristics of the current and the proposed practices, but also on the nature of the clinical problem (243,643,937). The dynamics of decisionmaking for preventive practices differ from therapeutic decisionmaking (240), and deciding which oral antibiotic to prescribe is different from deciding which diagnostic test to use. Decisions that can be reduced to explicit and clear algorithms may be more easily influenced by new evidence than decisions that involve a delicate balancing of numerous considerations.

**Physician characteristics.** The individual characteristics of a physician+. g., technical skills. demographic characteristics, attitudes. and other personality features-an affect the physician's openness to new information and his or her ability and willingness to change practice. Some group characteristics are also influential. For ex - ample. residents may be more receptive than other physicians to new information (21 4.679). Generalists and specialists have been shown to differ in their illfc~rnlation-seeking practices and in their attitudes about different types of evidence (240). Different styles of medical training can also influence practice style (226).

Practice setting. Characteristics of a physician's practice setting can influence the degree to which he or she responds to information (240.71 3). Such characteristics include the number and type of colleagues (e.g., solo vs. group practice; single vs. multiple specialty; office-bascd or institutional). degree of interaction among colleagues, involvement of physicians in medical education, characteristics of the patient

## FIGURE 8-2: A Coordinated Model of Implementation of Clinical Evidence




SOURCE J Lomas, "Retailing Research Increasing the Role of Evidence in Clinical Services for Childbirth," *Milbank Quarterly* 71 (3) 439-475, 1993 Reprinted with permission

population, workload, organizational culture, and practice styles of colleagues.

- **Environmental incentives and constraints. Environmental factors shown to influence physician practice and physicians' choice of therapy include regulatory influences (e.g.,** drug approval), insurance benefit coverage and payment policies, availability of resources, quality assurance and utilization review activities. and legal considerations. Government regulation and other environmental influences affect institutional decisions to acquire facilities and technologies, and those choices can in turn influence clinical decisions.

- **Credibility of information source. Physician decisions are strongly affected by the beliefs and practices of their clinical colleagues. Physicians are more likely to respond to information if it comes from** sources they know and respect, or from organizations with which they are associated (240,265,33 1,768). In contrast, physicians often reject information or directives that come from sources they perceive as antagonistic, such as health insurance companies or government entities (57,73,265.29 1). The adoption of a new practices by locally or nationally credible individuals (or organizations) may accelerate its diffusion in the medical community (306,466).

- **Channel of communication. The** source and mechanism by which messages are transferred affect their reception (240). Social networks and face-to-face communication with respected peers are particularly influential (102,69 1).

Local networks of communication are particularly important vehicles for transmitting information concerning new medical practices (306). Physicians may rely on these local networks because of uncertainties involved in adapting new clinical management strategies to their own patient population--either because of the particular characteristics of their patients or the resource constraints under which they practice. They may feel more comfortable adopting new recommendations if they know that their colleagues are doing the same.

- **Decisionmaking process. Decisions made by single parties can be changed more easily than those requiring multiple participation, particularly if institutional decisions are required (243). On the other hand, institutional decisions are increasingly** responsible for promoting changes in practice among groups of individuals, and they may foster a greater magnitude of change after a somewhat longer initial delay.

- Evaluative methods and evidence. The quali-(y of the evidence supporting a change in behavior may have an impact on the extent to which information affects physician practice. The quality of evidence may exert its influence via the discrimination of individual physicians, guidelines developers, or other institutions that review evidence and attempt to infuse it into practice. Even the most rigorous evidence. however, does not necessarily affect practice directly (box 8-1 ).

- **Patient factors. The patient's role in clinical decisionmaking has increased substantially over the past several decades. Patient preferences are important and legitimate determinants of the** course of diagnosis and treatment of medical conditions (41 9).

## APPROACHES TO CHANGING CLINICAL PRACTICE

Approaches to changing clinical practice *range* from passive dissemination of research results or other information to more active strategies to incorporate those results into practice. Simple active approaches employ a single type of incentive--e.g., an economic inducement-to promote change. More complex approaches employ a variety of different incentives in combination. This section reviews selected evidence of the effective - ness of various approaches, especially as they relate to the implementation of clinical practice guidelines.

### Passive Dissemination

The simplest approach to changing clinical practice is through the passive dissemination of in-

## BOX 8-1: Clinical Trials and Changes in Clinical Practice

Researchers have noted numerous examples of cases in which the results of large, well-designed studies did not appear to influence clinical practice (27) In a review of 28 studies that examined the impact on practice of specific clinical trials, nine showed that practices changed in the direction suggested by the trial. Only two, however, provided reasonable evidence that practice changed as a consequence of trial findings (240) Many Instances of new therapies that have been rapidly adopted in the absence of any clinical studies proving their value have also been documented (240),

Trials that have had an impact seem to share some common features Often, these trials study conditions and treatments for which the most relevant endpoint is death, an unambiguous negative outcome, In addition, most involved a discrete pharmacological intervention that was easily reproducible at other sites For example, in the case of the use of aspirin to prevent second heart attack, the intervention is inexpensive, simple to administer, and safe; and the avoidable outcome is severe. Affecting physician practice may be some what difficult when the results of trials in which the potential negative outcomes are less serious, less observable, and a broader range of therapeutic interventions are studied.

SOURCE Off Ice of Technology Assessment, 1994 based on sources as shown Full citations are at the end of the report

formation. Publication of a study's results in a medical journal, for example, is a long-standing mechanism for information dissemination. The appeal of this strategy is that it is generally relatively inexpensive, straightforward, and can make use of mechanisms already in place. However, the more complex models of physician behavior predict that its impact would be limited, and experience generally supports this prediction.

Studies of the impact of large, well-publicized randomized controlled trials (RCTs), which are hailed as the "gold standard" of clinical evaluation methods, on clinical practice illustrates both the possibilities and the limitations of the passive dissemination approach. In some cases, the publication of large RCTs has indeed had a strong influence on clinical practices. For example, the percentage of physicians reporting routine use of thrombolytic therapy after heart attack rose from 3 to *68* percent after the release of a major randomized study showing its benefits ( 137).[2] Changes of similar magnitude have been noted in response to large, well designed trials involving the use of aspirin for patients who have suffered a heart attack (438).

But passive dissemination is not dependable, even as a method of implementing high-quality evidence. The dissemination of the results of laser treatments for diabetic retinopathy (an eye disease common in persons with diabetes) provide a compelling example. In 1976, the first of a series of clinical trials was reported, showing that laser treatment for diabetic retinopathy was highly effective in preventing blindness if performed during the early stages of the disease ( 163). **The results were cited over 70 times in the medical literature between 1976 and 1979 (240). However,**

[2]Interestingly, a very large trial that reported similar benefits from thromboly tic therapy one year prior to the first surv ey had not had any effect on these physicians. 1( is possible that the results of this first study, though they failed to promote the recommended change on their own, helped predispose physicians to change once the results were corroborated by the later study.

18 months after the results were first reported. only one-third of internists and family practitioners were familiar with them. and most of those had learned about them from a colleague (737). In addition, most physicians were found to be unable to properly identify the condition itself during routine ophthalmoscopic examination (74 1 }—an inability which may have led them to ignore the trial results. Underdiagnosis and undertreatment of this condition are still widespread (65.79).

Whether or not mere disseminate ion of information is capable of influencing practice, the information itself does often serve a critical supportive function. Evidence from clinical evaluations can predispose or enable physicians to follow new practices or abandon inappropriate ones. It can also reinform the continued use of appropriate practices (298).

## Active Approaches

More active approaches to changing behavior have relied heavily on three particular sources of influence on clinical behavior: the economic and administrative environment: the influence of patients: and the influence of clinical peers. Several basic strategies have evolved that take advantage of these influences. In recent years. more integrated strategies for changing physician behavior have also emerged, combining elements that have been found to be effective in motivating change.

### *Economic Approaches*

Most analysts agree that the influence of economic incentives on clinical practice can be substantial (1 ,344,682). A number of studies, for example, have linked physician ownership of facilities with higher rates of patient use at those facilities for services such as diagnostic imaging (345), radiation therapy (534), physical rehabilitation. and psychiatric testing (743 ). In contrast, when faced with strict financial constraints for a particular service, physicians reduce the amount of that service that they provide (345 ). The rates at which physicians hospitalize patients or see them as outpatients have also been found to be associated with reimbursement (344).

Strong economic incentives can affect practice even when the incentives are targeted to institutions. For example. implementation of prospective, diagnosis-based payment for inpatient services under Medicare in 1984 resulted in shorter lengths of stay for patients with many diagnoses and, consequently. decreases in Medicare spending for inpatient care (706).

Although powerful. economic incentives have clear limitations, and they can be overshadowed by other factors. For example. recent studies of Cesarean delivery rates in managed care settings showed high rates of procedures considered to be inappropriate, even though no economic motive could account for these procedures (55 ). Another study found that reducing the insurance payment for Cesarean delivery to make the payment equivalent to that for vaginal delivery did not alter the Cesarean delivery rate (597).

Given that other factors appear to have equal or greater influence on physician behavior, strategies that use only economic incentives may be crude and inefficient mechanisms for changing physician practices. They may also sometimes produce undesirable results (e.g., by reducing the use of valuable as well as wasteful services}. In the RAND Health Insurance Experiment,[3] which randomized consumers to insurance with differing levels of cost-sharing, patients received up to 50 percent more care when the care was free to them (489). The additional serviecs were not necessari - ly the most valuable ones. however: payment incentives affected the rates of appropriate and inappropriate services equally, in both outpatient and inpatient settings (464,740). Likewise. studies of geographic variation in the use of high-cost procedures have found that areas that have a higher frequency of a procedure do not necessarily have a higher proportion of inappropriately performmed procedures. These findings suggest that

[3] The RAND Health Insurance Experiment was conducted between 1074 and1982.

broad economic incentives might not succeed in selectively reducing inappropriate use (115,118, 444,909).

Finally, financial incentives can sometimes have powerful unintended effects. For example, implementation of prospective payment to control Medicare spending on inpatient care spurred concomitant increases in outpatient care spending (61 7,756). In addition, many hospitals responded to reductions in Medicare income by raising their charges to patients with private insurance (616).

### *Administrative Approaches*

Physicians make decisions within the context of the institutions with which they work. Purposeful structuring of administrative rules, policies, and procedures thus can be a significant force for changing practices. Institutions implement such strategies for various purposes, including cost containment, quality assurance. and risk management (535).

A common administrative strategy to constrain the use of certain products is the use of prescription drug formularies. Requiring the use of special forms or other additional steps to obtain approval for nonformulary prescription drugs has been shown to significantly alter physicians' prescribing patterns (93,275).

Administrative strategies can also increase adherence to desired practices by eliminating hurdles or by facilitating a process. For example, in one study, a policy allowing nurses to administer influenza vaccines without physician orders increased the frequency of vaccinations (491). Computerized reminders have also been used successfully to change physician practice (e.g.. computerized dosing protocols to improve the accuracy of dosing for potentially toxic drugs (401)).

Utilization review (UR) is an administrative strategy used frequently by third-party payers to contain costs and monitor the quality of care. Traditional UR involves case-specific review of the appropriateness of care being contemplated or provided and includes techniques such as prior review, preadmission review, continued stay review, second surgical opinions, and case management of high-cost cases (369). All of these techniques rely on the creation of a parallel decisionmaking entity that independently assesses the appropriateness of an intended clinical action.

Empirical evidence on the effect of these traditional UR programs is mixed. Some studies show that payers who use UR programs have lower rates of hospital use than those who do not (231). However, some of these savings may be offset by increases in the use of outpatient services. Little evidence exists to suggest that long-term patterns of expenditure growth are altered by adoption of UR methods (369). A comprehensive review of second surgical opinion programs concluded that the impact of these programs has not been adequately demonstrated (460). Few studies have addressed the impact of UR programs on quality of or satisfaction with care.

In the past few years, UR has evolved from a narrow focus on length and necessity of hospital stays to more detailed assessments of the clinical appropriateness of the care provided. Several commercial UR firms now provide employers and insurers with detailed reviews of the appropriateness of anticipated clinical procedures. Experience with this more detailed level of review is still limited. Anecdotal evidence suggests that it may reduce procedure rates in some cases, but it is possible that some of this reduction comes from patients who are not referred for surgery even though the surgery would be considered appropriate for them.

Administrative and economic interventions may be particularly useful when desired practices can be described simply, are well supported by clinical studies, and are unlikely to provoke significant physician or patient resistance. Like economic incentives to promote practice change, however, administrative strategies can have unintended consequences if applied without appropriate quality controls. For example, limits on Medicaid prescription drug reimbursement were associated with increased hospital and nursing home admissions among Medicaid-dependent

elderly patients (714). The increased admission rates not only reflected a compromise in the quality of care, but probably also more than offset any savings generated by the prescription drug limits (714).

Physicians may seek to circumvent restrictions that they believe are unjustified or externally imposed (57). If the restrictions do in fact lead to compromised quality of care, such circumvention may be desirable. However, if physicians rebel against medically justified recommendations, quality of care may also be compromised and the administrative strategy can become costly to enforce.

Both administrative and economic incentives—even those that are effective in producing short term changes in behavior—may need to be continually applied to prevent a return to previous practices. The lack of durability of change associated with these externally applied strategies has been documented in association with computerized presentation of laboratory test costs (760) and telephone reminders to encourage early discharge for some diagnoses (902).

### *Approaches Using Patient Influences*

Most strategies to change clinical practice are aimed at physicians. But the idea that physicians can be unilateral decisionmakers, divining patients' preferences and incorporating them into a clinical decision, has been increasingly challenged, and the role of patients in decisionmaking is becoming more prominent (419,698).

Patient involvement in clinical decisionmaking is a legal requirement as well. All 50 states have statutory requirements for informed consent in medical care (615). Although the required elements of disclosure and specific legal formulation vary from state to state, the basic concept of informed consent is that patients must be informed of the risks and benefits of alternatives therapies before undergoing treatment by a physician.

Patients are participants in clinical decisionmaking and clearly affect the choice of medical therapies. Recognizing this, pharmaceutical companies and other commercial interests now frequently market their products directly to patients. High demand for many experimental therapies, such as autologous bone marrow transplant for metastatic breast cancer (see chapter 5), and experimental drugs for other cancers and AIDS (acquired immunodeficiency syndrome) (787) are examples of how patient demands can influence clinical practice.

Patient education has frequently been used to promote certain disease prevention practices and long-term management of chronic conditions. For example:

- Compliance with cancer screening recommendations has been improved through a program that combined patient and physician interventions (516).
- A program that encouraged diabetic patients' participation in their care improved both control of blood sugar and overall quality of life (302).
- The functional status of patients with arthritis has been improved by educating patients concerning what to ask their doctors during clinic visits (415).
- Use of prostate screening tests has been reduced by preparing physicians to respond to patient inquiries on the topic (738).

Several groups have developed standardized methods for incorporating patient preferences into clinical decisions. The most sophisticated system to date uses video disc technology to present information on risks and benefits of alternative choices in therapy. The Foundation for Informed Medical Decisionmaking has developed video discs addressing four common medical problems: benign prostate disease, high blood pressure, low back pain, and early stage breast cancer (940). The best studied of these applications is benign prostate disease (see box 8-2).

The importance of patient preferences and the impact of soliciting those preferences varies with the clinical condition involved. For example, choice of treatment for benign prostate disease involves balancing the risks and benefits of various

## BOX 8-2: Use of Video Discs To Inform Patients of the Risks and Benefits of Alternative Treatments for Prostate Disease

The Foundation for Informed Medical Decisionmaking, a nonprofit organization, has developed video discs to facilitate informed decisionmaking by patients for four common medical problems: benign prostate disease, high blood pressure, low back pain, and early-stage breast cancer (940) Benign prostate disease is the best studied of the video disc applications

The disease was chosen in part because research showed significant complication rates for men who had undergone surgery, and there was concern that the risks were not being reliably communicated to patients by their surgeons (918) The primary benefits of surgery for this disease are reduced frequency of urination and increased force of the urinary stream; risks include impotence, retrograde ejaculation, and urinary incontinence Research also suggested that the level of concern about particular risks varied considerably across patients, and could not be predicted by any objective measures (e.g , clinical history, physical findings, urine flow, or even level of symptoms) (918)

The video disc program discusses on the risks and benefits of alternative treatments for benign prostate disease The Information is presented by clinicians and by patients who have undergone the treatments and experienced more and less desirable outcomes. Use of the video was expected to decrease surgery rates

Results of a cohort study begun in 1990 suggest that patients who watched the video disc frequently choose conservative management (i.e., watching and waiting) over surgery (548,918) In two prepaid group practices that participated in the cohort study, application of the video disc has also reduced overall surgery rates for benign prostate disease (91 8). An ongoing research effort is examining the extent the changes noted in patient preferences and decline in surgery rates can be attributed to the video disc intervention (41). Other factors, such as a general trend towards conservative alternatives to surgery and the availability of new drugs to treat the symptoms of benign prostate disease, may explain part of the reduction in surgery rates (41).

SOURCE Office of Technology Assessment, 1994, based on sources as shown Full citations are at the end of the report

functional impairments. However, some patients with chronic or potentially fatal illnesses may prefer aggressive treatment even if it promises only a small probability of benefit. In a recent example, demand for a newly approved drug for multiple sclerosis has substantially exceeded supply, despite absence of evidence of long-term benefit (455). On the other hand, other patients with known terminal conditions may prefer less intensive intervention at the end of life.

Improving patient involvement in decisionmaking is fundamentally an ethical goal, driven also by legal considerations. In some cases, patient involvement may facilitate reductions in inappropriate care, potentially reducing the costs of care. In others, however, it will increase the use of appropriate or desired services, leading to improved health for patients (but not always lower costs). For example, while the video discs for low back pain and benign prostate disease might be expected to reduce the use of surgery, similar interventions for high blood pressure and early stage breast cancer might be expected to increase the use of certain hypertensive drugs and chemotherapies (699). Patient education and communication might also increase appropriate resource use con-

siderably when applied to populations that have been historically underserved.

Thus, strategies that rely on educating patients and increasing their involvement in clinical decisionmaking should improve people's health and satisfaction with care. Policy makers cannot necessarily expect these strategies to reduce costs, however.

### *Approaches Using Clinical Peer Influences*

The influence of experts and local information networks (see p. 5) is illustrated by the effectiveness of "opinion leaders" in causing changes in practice. The opinion leader strategy makes use of individuals within a local community who are regarded by their professional colleagues as authori - tative sources for information and advice. The success of the strategy thus depends on the willingness of the opinion leaders to both accept and promote the practice in question. Opinion leaders have been successful in improving clinical management of patients with arthritis (736), reducing rates of Cesarean delivery (470,527) and reducing length of stay for intensive care unit patients (193).

A related form of peer influence strategy, "detailing, " was first developed and used for commercial purposes. For years, drug companies have capitalized on the credibility y of local sources of information and communication networks, using informal personal contacts between drug company representatives and physicians to market their products (31,1 02,470,712,736,937). The essential features of detailing programs are one-on-one interactions between the educator and the physician, precise assessment of information needs, use of a credible source, carefully crafted messages, and attractive informational materials (31,7 12). The same principles have been successfully incorporated into educational interventions such as promoting the use of safer, less expensive drugs; encouraging more appropriate dosing intervals; and reducing inappropriate blood transfusions by surgeons (32. 153,296,467,536,712,7 16). The effectiveness of detailing has been demonstrated in both outpatient and inpatient settings (710,7 12).

Not all detailing efforts have been successful. In one failed effort to reduce use of antipsychotic drugs in nursing home patients (627), it was suggested that educators were not credible and that the educational visit was too brief and involved limited interaction (536). Detailing programs are labor-intensive and require development of materials tailored to specific target populations, making this type of intervention potentially costly. These approaches may, however. provide savings greater than their cost when directed to reduction of use of wasteful or expensive therapies (711),

### *Practice Profiling and Feedback*

Practice profiling, information feedback to physicians, and other techniques that combine data gathering with peer influence are approaches being increasingly used with the goal of reducing variation in the rates with which physicians perform specific procedures. Physicians in Maine, for example, have successful y used profiling and feedback to reduce the variation in the rates of hys - terectomy and prostatectomy (see box 8-3).

Practice profiling involves collecting data describing the practice patterns and outcomes of clinicians and feeding these data back to them so they can compare their own practice patterns to those of their colleagues or, in some cases, to an objective "benchmark" (441 ). Activities commonly measured in practice profiles include rate of use of selected services (e. g., screening tests), clinical outcomes for certain types of services (e.g., proportion of cardiac bypass surgery patients who die within 30 days of surgery), and costs of care. Common uses of profile data include educational feedback to physicians. utilization review, and coverage decisions by insurers.

Reviews of the literature on use of physician profile data as an educational feedback tool have found it to be an effective means of changing prac - tice ( 153.296,536,543,677). The technique has been particularly useful for improving prescribing practices, increasing use of preventive services, and decreasing use of laboratory or x-ray tests. Feedback interventions designed to increased use are especially effective, while those designed to

### BOX 8-3: The Use of Practice Profile and Feedback by the Maine Medical Assessment Foundation

An ongoing effort by the Maine Medical Assessment Foundation (MMAF) uses profile data to reduce regional variation in the rates of selected procedures within the state. MMAF emerged from a collaboration of physician health services researchers, the Maine Medical Association, and Blue Cross and Blue Shield of Maine

MMAF convenes specialty study groups to focus on a number of medical conditions and surgical procedures (483), Each study group meets periodically to review data on utilization rates of certain practices. Groups are headed by a physician leader selected for prominence within the relevant specialty When regional variations are noted, the group examines potential reasons for the variability. The groups do not develop specific protocols or indications for practice, nor do they attempt to determine optimal procedure rates (483),

When significant variation is noted, MMAF takes its data to physicians in high rate areas and sponsors meetings among those physicians to discuss reasons for the variation Although the variations discussed are compiled as regional rates, the small number of providers from each region means that participants are usually able to determine which geographic rates apply to them, thereby providing "individual" feedback (423), Physicians have been surprised and uncomfortable to learn of the degree to which their own practices may not be typical and are motivated to bring their practice patterns more into greater conformity with that of their peers (1 14,423), The intervention thus takes advantage of the importance of locally defined community practice standards, as well as the use of opinion leaders, who participate as leaders in each study group.

This "consciousness-raising" approach has reportedly reduced the variability in practice in treatment of prostate disease, hysterectomy, back pain, and pediatric hospitalizations—most often by reducing the rates in areas of Maine with high utilization.

For example, rates for hysterectomy declined precipitously in the region noted to have the highest rates within a single year after data were presented to the area physicians at a group meeting (483) The meeting was attended by representatives of the state medical association and BCBS,

decrease use are somewhat less so (677). The strategy is most successful when it is individualized, delivered by a respected peer, and involves comparisons with the performance of peers (214).

The success of feedback interventions may depend on the fact that they are usually tested in settings where clinicians have already agreed to the process of review and response (424,483,543, 920). Strategies that provide feedback passively to clinicians have been less successful than programs in which feedback was preceded by some discussion or standard-setting exercise. Physicians must also be convinced that the proposed new practice is in fact of equal or greater benefit to patients than the old practice. In a study of the use of computerized feedback to lower prescription drug costs by encouraging residents to substitute lower cost alternative medications, for example, no effect was observed because the physicians were not convinced the substitute medication was as effective (543).

Profile information has been most useful when it is delivered in a form that can be translated into a specific improvement effort (677). Informing a hospital that it has a higher than average death rate from pneumonia provides little information about the causes or solutions for that problem. On the other hand, profile information that relates surgi-

**BOX 8-3 continued: The Use of Practice Profile and Feedback by the Maine Medical Assessment Foundation**

as well as by two professors of obstetrics and gynecology from the University of Portland These well known and respected clinicians had previously been involved in training several of the younger practitioners from the high volume area (1 14) Multiple factors probably contributed to the observed change in practice, Including the educational value of the profile data, the desire of physicians not to be practice outliers, and the potent effect of personal mentors who were expert clinicians.

In the case of variation in use of back surgery for management of lumbar disc disease, the MMAF-sponsored group discussion led to recognition that there was very little objective evidence to provide guidance on selecting patients who would be most likely to benefit from surgery. Despite this lack of reliable guidance, surgeons with high rates did subsequently lower their rates, even though no explicit changes were made in patient selection criteria (1 14). It is widely believed that surgery is overused Reduced use of surgery for back pain is probably an Improvement in care, since many disc problems are resolved with bed rest However, in the absence of valid patient selection criteria, it is possible that physicians are reducing their rates while still providing surgery to the wrong patients In fact, despite continued meetings of orthopedic study groups, rates of disc surgery in Maine have begun to rise again, a phenomenon that may be due in part to the absence of convincing evidence on the relative benefits of surgical vs. conservative strategies (424)

For other services, such as cardiac catheterization, the process has not had clear effects (483) For this and other services, however Interpreting data on the effects of MMAF feedback activity can be difficult, because they do not account for underlying trends in utilization rates that might not be associated with the group Intervention.[1]

[1] To help address this prob'em AHCPR has funded a five-year, $32 mi Ill on dollar study 10 look at the impact on practice of these study group methods in Maine New Hampshire and Vermont

SOURCE Office of Technology Assessment 1994 based on sources as shown Fu'1 citations are at the end of the report

cal wound infection rates to the timing of antibiotic prophylaxis administration gives better guidance on how to improve care. However, in order to provide information in this more useful form, profile data developers may need access to data other than what is routinely available in an administrative database. Practice improvements using profile data are therefore likely to require planning. effort, and expense.

An implicit assumption of profiling efforts that attempt to bring "outlier" physicians closer to the average practices of their colleagues for certain procedures is that "average" practice is appropriate. This is not always the case. For example, profiles of current practice would suggest that over 20 percent of women should have Cesarean deliveries—a suggestion strongly contradicted by available research evidence.

Furthermore, research has shown that physician practices vary considerably across both institutions and geographic areas (see chapter 2). The average practice in one institution could be appreciably different from the average practice in another.

Maine's experience with profiling feedback through peer group discussion (described in box 8-3) shows that it is possible to change practice without generating any new information about the relative effectiveness of alternative management

strategies. By studying those practices that are suspected of being overused, or that have recently increased in use without explanation, the assumption is generally made that high utilization rates are inappropriate. As illustrated by the experience with feedback on treatment for back pain, however, when objective evidence of effectiveness is limited, the group model may be less effective.

Incorporating evidence-based benchmarks of appropriate practice, rather than averages, as the standard for comparison improves the likelihood that practice will actually move towards better quality care. Where strong evidence is not available, profiling efforts can still attempt to analyze data in ways that suggest appropriate paths for change (e.g., by examining patient outcomes), rather than simply to encourage a move towards "average" practice.

### *Combining Feedback and Economic Incentives*

Insurers are increasingly using profile data rather than detailed review of individual cases for utilization review. A number of anecdotal reports suggest that physician practices may change considerably when payers provide information showing them that their rates of use of specific services are significantly above normal range. One payer, for example, found that clinicians reduced rates of colonoscopy when informed their rates were unusually elevated (677). The ability of the payer to deny payment, or institute other sanctions, may reinforce whatever educational value the profile data offer. Similar threats of payment denial based on outlier status have been effective at shortening lengths of stay and reducing the prices charged for specific services (677).

Publication of cardiac surgery mortality data in New York State offers a more subtle example of how financial and social incentives can reinforce feedback of profile data to influence practice. For the past several years, the New York State Health Department has required all hospitals that perform cardiac bypass surgery to submit a standardized data form on their patients. From these data, risk-adjusted mortality rates for each hospital have been developed. These data have revealed significant variation in the death rates among the hospitals. The state originally intended to provide the data confidentially to hospitals and surgeons, but as the result of a lawsuit, the data were released publicly in a newspaper article (739).

In the two years since initial release of the data, adjusted mortality rates for cardiac surgery have fallen in the state, despite increases in the rates of cardiac surgery and in the average severity of illness of patients undergoing surgery (447). This trend is consistent with the possibility that hospitals and surgeons took actions that improved their surgical procedures, and there is evidence that some at least attempted to do so. One hospital, for instance, temporarily closed its cardiac surgery unit while investigating potential causes for poor outcomes. Several hospitals sought the assistance of the health department in determining the cause of poor outcomes (739).

The mortality profiling strategy relied in part on the social influence of having one own clinical practices compared with the practices of one's peers. Data derived from this strategy also raised the possibility that physicians' care was resulting in avoidable patient deaths, a severe outcome that served as a strong motivator to identify and correct inadequate practices. Moreover, public release of the data raised the possibility of lost prestige and income, which reinforced the motivation to change practice despite uncertainty regarding the validity of the information.

This example illustrates some of the dangers of using provider profiles alone as quality indicators. The validity of the profile information was highly sensitive to both the accuracy of the data provided on each physician and the at tent ion to other factors that could have influenced rate differences. For example, it has been suggested that the apparent improvement in mortality after release of the data could be due to increased documentation of patient severity of illness, which could have artificially reduced the risk-adjusted mortality rates (362). It has also been suggested that physicians may have become more reluctant to operate on sicker patients for fear of unfavorable outcomes-

an unintended consequence that would not only lead to decreased surgical mortality, but potentially hamper access to surgery for patients with the greatest need. Interestingly. despite the considerable controversy concerning public release of the data, there is no evidence that patients' choices of surgeons have so far been affected by the information.

### *Continuous Quality Improvement*

Continuous quality improvement (CQI) systems, the formal application of quality improvement techniques, are being increasingly applied to clinical practice (57,68. 157.326 .402,571,692.703,7 19).

In essence. CQI is an integrated administrative strategy that involves three phases: problem identification, clarification of the sources of the problem, and design of a strategy to resolve the problem. Clinical practice profile data are often used in CQI to identify areas where performance problems may exist, usually to help understand why the problem is occurring, and to track changes in results when actions are taken to resolve the problem (677), CQI methods have also been used to promote implementation of clinical research findings (93) and practice guidelines (376).

The conceptual basis for CQI is that it is more effective to improve the average performance of all members of a system than to identify and penalize the small fraction with the worst performance (57). Physicians who vaccinate too few of their patients could be singled out and penalized in order to provide an incentive for them to improve, However, it may be that tight scheduling of patients makes it easy to forget to order vaccinations, and that vaccination rates could be improved overall by using a simple computer tracking system that provides reminders or automatic orders (57).

While few CQI efforts have generated sufficient data to document an impact, there is a broad belief that such initiatives can improve clinical outcomes, increase patient and provider satisfaction. and even reduce costs of care. Among the best known examples of the successful application of CQI methods to change practice are the efforts of the Harvard Community Health Plan (HCHP) and Intermountain Health Systems. The experiences of these providers, described briefly in box 8-4, illustrate the significant improvements that can be achieved through application of CQI methods, but they also show that a CQI system can be time-consuming to design and implement and that its success may depend heavily on the availability y of computerized support systems. Application of CQI techniques may be less successful for services where disagreement about risks and benefits exists, or that involve more complex legal or economic factors.

Also important to note is that HCHP's successful improvement in screening for cervical cancer (box 8-4) relied on increasing the use of a service that is viewed as desirable by physicians and patients. CQI methods may not be as successful in reducing the use of services that some physicians still feel arc appropriate. An effort at HCHP to reduce the inappropriate use of x-rays and endoscopy for evaluation of stomach pain was not successful in changing practices. even among the physicians who participated in the team that developed the protocols (301 ).

CQI methods do incorporate important features known to be associated with change and learning. They are noncoercive and typically encourage active participation by local clinicians. The gradual nature of the process, and the emphasis on exploration of the causes of problems, helps to ensure that new information is fully understood by those who are expected to change their behavior. The available experience suggests that where institutional commitment is high, time and resources are adequate, and data systems are available, CQI may result in improved processes of care.

## Lessons from the Literature

The previous section reviewed a variety of strategies that have been used to change physician behavior. Unfortunately, the literature on physician behavior change has four serious limitations that make it difficult to evaluate the relative effectiveness of different strategies.

## BOX 8-4: Using CQI to Promote Practice Change: Two Examples

The Harvard Community Health Plan (HCHP), a large staff-model health maintenance organization in Massachusetts, has successfully used continuous quality improvement (CQI) methods for a number of years to change medical practice (676), One example involved an attempt to improve the quality of the cervical cancer screening program by ensuring that screening was performed on all eligible patients, specimens were adequately collected by physicians, and appropriate responses to results of the test were undertaken To Improve the adequacy of specimens collected, all physicians were trained in use of a cervical brush, which was added to all cervical smear collection kits Use of this brush was known to dramatically improve the quality of specimens A working group then reviewed clinical studies and developed an algorithm for the clinical actions indicated by each possible smear result, The actions were printed on the laboratory result form A computerized patient information system also automatically checked that the recommended action was completed—a step that markedly increased adherence to the protocol. After SIX months, 99 percent of patients receiving smears had received care indicated by the algorithm (676)

In a few cases, CQI efforts have actually revolved identifying the relative effectiveness of different care practices and implementing the more effective ones more broadly One of the best known examples of this use of CQI is an effort by Intermountain Health Systems, a multihospltal system in Salt Lake City, Utah, which has a highly developed medical information system that IS being used to track data on hundreds of different patient treatments and outcomes (1 25),

Intermountain examined computerized data to compare rates of surgical wound infections for nearly 3,000 patients undergoing elective surgery The data revealed that patients who received antibiotics two to 24 hours prior to surgery had much lower rates of post-operative infection than those who received antibiotics within two hours before surgery or within three hours after surgery (124) Since the study, 96 percent of surgeons at Intermountain are giving antibiotics more than two hours before surgery, and the surgical wound infection rate has dropped from 18 percent to O 4 percent, resulting in significant cost savings for Intermountain hospitals (493),

Practice change in this case may have been enhanced by several factors, First, almost all surgeons were using antibiotics, and the new practice only involved a small change in the timing of that therapy, Second, the practice was simple to undertake—no additional time or equipment was necessary Third, the data were undeniably applicable to the surgeons involved, since the data were drawn from their own patients Fourth, the magnitude of the difference in infection rates was large enough to make the result both credible and perceived to be of clinical significance Surgeons pay attention to infection rates in their patients, since for the most part post-operative infections are considered to be avoidable and are thus a reflection on surgical technique and competence The success of Intermountain initiatives may not be readily generalizable to other settings due to this provider's unusually high degree of record computerization (1 25)

SOURCE Office of Technology Assessment 1994, based on sources as shown Full citations are at the end of the report

First, efforts to evaluate behavior change strategies have not been particularly rigorous. They typically rely on quasi-experimental methods and, sometimes, on purely anecdotal information. A recent review of 777 articles describing educational interventions found that only 86 were conducted as randomized trials and only 50 provided enough data to be part of the researchers overview ( 153). In another review of behavior change literature, the reviewers demonstrated that every well-controlled study of one strategy was negative, whereas all uncontrolled studies for the same strategy showed a positive impact (7 12).

A second, related problem is that many studies do not address all relevant aspects of the clinical problems involved, the procedure involved, or environmental characteristics such as the practice setting in which the strategy was implemented. Lack of attention to such factors makes difficult the determination of which aspects of the strategy accounted for its success or failure. and whether the results arc generalizable to other physician populations.

Third, most studies concentrate on the clinical practices that tend to be the easiest to measure: compliance with preventive care guidelines, use of laboratory testing, and drug prescribing (216). These services, while important, represent only a small portion of the range of health care decisions (679). Because the bulk of research on practice change has focused on only a few types of decisions, it is difficult to generalize findings about successful strategies to more complex clinical decisions, such as choosing between medical or surgical treatments.

Finally. most studies used resident physicians and medical students as subjects. Students and residents are generally more motivated information seekers and probably more impressionable than arc other physicians. Thus, many strategies that prove effective in these populations may not be as effective at changing the behavior of more experienced physicians.

Despite these limitations. the existing literature does reveal some consistent themes. Recent detailed reviews of the literature ( 153,296) have concluded that educational interventions were most likely to be successful if they had the following characteristics:

- Opportunist y for direct participant ion and personal interaction (296). In many cases, substantive changes are not required. on] y a sense of having put a personal mark on the product (290.376).
- Strong support for the practice change among community physicians (296).
- Smaller group settings (153).
- Individual instruction and feedback (153).
- Use of respected clinical leaders(116,382,488).
- Identifying the specific learning objectives of participants ( 1 53).
- Higher specificity in both the subject matter addressed in an educational program and in the audience targeted ( 153.923).
- Flexibility to modify and adapt innovations to local circumstances (731).
- More intense interventions (153).

The literature also suggests quite strongly that combinations of interventions are more likely to be effective than single intervention strategies (153,296,467,). Strategies that combine educational and administrative interventions may be particularly effective in altering clinical practices (467 ). Educational strategies themselves are more likely to be successful if they are interactive and personalized, delivered by respected peers, and coupled with practice profiling and structured feedback.

The success of strategies to change practice also varies depending on the type of clinical task involved. Davis and colleagues ( 153) reviewed studies evaluating strategies to change physician practice in the areas of test ordering, drug prescribing, use of preventive services. clinical management strategies, and patient counseling. Ten out of the 11 test-ordering interventions studied were successful. These interventions represented a range of strategies, including feedback, reminders, and chart review: providing cost information; and training in clinical reasoning. Five out of 6 drug prescribing interventions were found to be successful in altering practices, predominantly through use of academic detailing or feedback of

information on drug use to individual providers. Fewer of the interventions to modify clinical management and patient counseling were successful. When successful, these interventions were more likely to involve individual interaction with experienced clinicians and computer-generated reminders (153).

Practice setting is also clearly important to success. For example, Soumerai and colleagues (7 13) note that feedback and reminder systems may work best in group practice settings, while in other settings detailing is more potent. The increasing trend towards health maintenance organizations (HMOs) and group practice may bode well for future efforts to systematically promote changes and improvements in medical practice.

Medical school **curricula** increasingly are experimenting with means of improving clinical decisionmaking. This includes training that highlights the importance of patient preference in decisionmaking; enhancing skills in searching, critiquing and using the published literature (330); and greater sensitivity to the costs of diagnostic and treatment services (229,760,936).

Given the variety of circumstances that surround clinical medicine, there can be no single "magic bullet" strategy for changing clinical practice. Rather, interventions need to be targeted to the condition and practice considered, taking into account the physicians' characteristics, clinical setting, and other relevant factors.

## IMPLEMENTING CLINICAL PRACTICE GUIDELINES

Clinical practice guidelines are a potentially powerful tool for communicating the findings of clinical studies to practitioners. The current federal interest in sponsoring guidelines development is motivated in part by the expectation that guidelines will result in changes in clinical care, improving the quality of care and reducing use and costs of unnecessary care.

At present, most guideline development efforts at the federal level rely on educational approaches for this implementation. National Institutes of Health (NIH) consensus statements and AHCPR-sponsored guidelines, for example, are disseminated through direct mailings, and through publication in popular medical journals (box 8-5). The results of research sponsored by these agencies similarly are disseminated through journal publications or presentations at medical conferences.

In the near future, however, policy makers will not necessarily be content to rely on passive education to promote guidelines. More coercive administrative and financial strategies are being contemplated, and some are already being implemented. Medicare, for example, is developing utilization review strategies based on guidelines sponsored by AHCPR and others (533). One private insurer recently announced that coverage for a series of different clinical interventions would depend on whether clinicians had followed clinical practice guidelines (74). A number of policymakers and commentators have suggested that health insurance benefits under health reform should be based on evidence regarding effectiveness and clinical practice guidelines (see chapter 2). These strategies assume that educational and other relatively benign strategies alone will not be sufficient to change practice.

### Impact of Guidelines on Clinical Practice

The impact of guidelines is measured most often by changes in physician practices. Few studies have examined the impact of guidelines using patient outcomes. The literature that exists is also somewhat contradictory. Some reviews suggest that guidelines have had little impact on practice, while others report greater promise.

For example, an extensive review of the products of the Consensus Development Program at NIH concluded that physician practice is usually unaffected by the results of these recommendations (41 1,43 1). A recent review of 19 studies that evaluated the impact of these and other consensus statements on practice concluded that few instances of impact are identifiable (465). Substantial discrepancies between guideline recom-

## BOX 8-5: Guidelines Dissemination at AHCPR

In establishing the Agency for Health Care Policy and Research (AHCPR), Congress was particularly concerned that the results of the work of the agency be widely disseminated As a result, one divison of the agency is dedicated entirely to this task The primary method of disseminating AHCPR guidelines is direct mailing, based on established mailing lists as well as requests received through a toll-free telephone number.[1] The demand has been strong, with several million of the first eight guidelines having been malted, About half of requests have been from clinicians, with the other half coming from consumers and other interested individuals Demand has been stimulated by media coverage of the guidelines One large surge of requests for publications, for example, was stimulated by reference to a guideline on urinary incontinence in a column by Ann Landers Many of the guidelines have been excerpted and published in peer-reviewed medical journals as well To enhance the visibility and credibility of the guidelines it sponsors, AHCPR also actively seeks the endorsement of relevant professional societies, and it has received many endorsements for each guideline

While a number of anecdotes have been collected suggesting that AHCPR guidelines have been well received in certain cases, there is as yet no empirical evidence of their effects on cliniclans' or patients' awareness, beliefs, or practices. This is due in part to the relative newness of the program AHCPR has issued a request for proposals and will fund research on patterns of use related to the guidelines (822)

AHCPR sponsors some research in dissemination methods and has initiated several studies specifically to evaluate various strategies for implementation of its guidelines Most of these studies involve comparisons of the range of strategies discussed in the chapter (e g , opinion leaders, CQI, practice profiling) as approaches to altering physician behavior (805).

In order to facilitate use of guidelines in generating performance benchmarks for profiling care, AHCPR is also working on methods to translate guidelines into review criteria This Involves deriving a discrete set of steps though which the recommendations in the guideline are converted into "yes" and "no' questions, which can then be applied to the review of patient care Application of such criteria could be used for a variety of purposes, ranging from providing practice feedback to physicians to potential use by payers to make coverage decisions Review criteria based on well-founded guidelines would clearly be preferable to more arbitrary or less evidence-based review criteria. However, the potential use of guidelines-based review criteria by payers may affect the expert panels' willingness to issue explicit recommendations for practice, and it might undermine physicians' confidence in the guidelines as an educational tool.

[1] AHCPR guidelines are produced in three formats a book version of about 100 pages a clinician summary of about 20 pages and a consumer pamphlet of four to eight pages

SOURCE Office of Technology Assessment 1994 based on information from the U S Department of Health and Human Services Public Health Service Agency for Health Care Policy and Research, and other sources as shown Full citations are at the end of the report

mendations and actual practices have been documented for a variety of clinical situations, including use of screening mammograms (5 17, 556,943), diagnostic staging studies for cancer patients (26 1), dehydration in treatment of childhood diarrhea (60), and others. This cumulative evidence was interpreted by recent reviewers to indicate that "clinical practice guidelines have been remarkably unsuccessful in influencing physicians" ( 153).

Other reports of the influence of guidelines are considerably more positive. Decreased use of cardiac pacemakers (425) and increased routine use of monitors during anesthesia have been attributed to clinical practice guidelines (211 ,608). Other studies have shown guidelines to lead to reductions in overall lengths of hospital stay and reduced use of intensive care units (193,470, 902,903), more accurate dosing of therapeutic drugs (439,630), and improved health for diabetics patients (887). A recent systematic review of 59 evaluations of the effect of clinical guidelines on medical practice found that all but four of the studies reviewed identified significant changes in practice in the direction proposed by the guidelines (308). The magnitude of change varied considerably across the studies, however.

The conflicting conclusions of these two bodies of evidence concerning the impact of guidelines on clinical practice deserve closer attention. A review hereof selected guidelines development and dissemination efforts helps to illuminate some of the factors that underlie the mixed results.[4]

### *Preventive Care Guidelines*

The use of widely accepted screening tests for common cancers and other conditions has been well below expected levels, suggesting that the simple existence of recommendations from national guideline panels on preventive services is inadequate to alter practices (332,5 17,556,943). Increasing the rate of preventive screening even through more active measures is sometimes difficult. In one study, performance did not improve among resident physicians who were taught about the recommendations of guidelines in small groups, nor did any change occur when patient-specific recommendations were taped to the front of each patient chart (333).

Computer-generated reminders, however, have been successful in most of the studies of the use of this technique in improving use of preventive services (40,504,505,506,654,759 ). These reminder systems use patient age, gender, and risk-factor information to determine which preventive services are indicated for the patient.

Generalizations from these studies of preventive practices to guidelines involving other clinical tasks may not be warranted, however. Preventive care is generally a population-based practice that involves minimal tailoring for individual patients. Screening guidelines also differ from guidelines for many other services in that they present relatively discrete and explicit recommendations for practice.

### *Cardiac Pacemaker Implantation*

A marked decrease in the use of cardiac pacemakers occurred in the early 1980s. The decrease has been credited to the dissemination of guidelines issued by the American College of Cardiology (ACC) and the American Heart Association (AHA) (425). The impact of the ACC/AHA guidelines was probably not the result of simple dissemination alone, however. First, public attention to inappropriate use of cardiac pacemakers was high in the late 1970s and early 1980s, with several research studies and a Senate hearing contributing to the concern ( 120,304,588,791). Sec-

[4] Studies included in this review were obtained through a MEDLINE[R] search, and from existing reviews of guideline implementation, as well as bib] iographies of relevant articles and personal collections of experts ( 153,216,376,40 1,535.946). Only the reviews by Dav is et al. ( 153), Grimshaw and Russell (308) and Johnston et al. (40 1 ) provide an explicit description of the methods of I iterature search and study selection used in their over-views.

end, a number of health care payers, including Medicare, used the guidelines as a basis for prior authorization review of pacemaker implantation and denied authorization for patients not meeting the guideline criteria *(304).*

### *Cesarean Deliveries*

A rapid increase in rates of Cesarean delivery prompted national guidelines in both the United States and Canada that concluded that at least some reasons for performing the procedure were unjustified (492,744,841 ). Several studies documented that U.S. Cesarean rates did not decrease after publication of the NIH-sponsored guidelines (283,43 1). Similarly, in Canada, 90 percent of physicians were aware of the national guideline, and over 80 percent agreed with the recommendations, but rates of the procedure continued to increase following guideline dissemination (468). Frustrated by the lack of response to guidelines based on such a strong evidence base, researchers and institutions turned to more active implementation strategies.

In Canada, Lomas and colleagues carried out a randomized controlled trial comparing the impact of different strategies to reduce rates of Cesarean delivery for women who had undergone a previous Cesarean---a clinical decision about which the guideline was quite explicit (470). Strategies tested included simple passive dissemination. profiling and feedback, and the use of opinion leaders to influence clinicians. No changes in practice were noted in either the passive dissemination group or the profile and feedback group.[5] In the opinion leader group, however, there were substantial increases in trials of labor and vagina] deliveries for women with prior Cesareans, as well as a reduction in the overall Cesarean delivory rate, with no detected change in birth outcomes (470). The success of this intervention may be due in part to factors such as the strong base of evidence in support of the recommendations (106,324), the large amount of inappropriate use suspected, and the clarity and simplicity of the recommendation involved. However, despite these favorable characteristics, an active implementation strategy was still required.

In the United States, an active intervention at Mt. Sinai hospital in Chicago reduced the rates of Cesarean delivery from 17.5 to 11.5 percent over a two-year period (55 1 ). Explicit policies were developed to encourage a trial of labor for women with previous Cesarean deliveries, and the policies were provided to both physicians and patients. Physicians were required to obtain and document a second opinion prior to performing a Cesarean delivery (other than emergency procedures). Data on each physician's rate of Cesarean delivery were available so that physicians could compare their own performance with that of their colleagues. Cesarean deliveries not complying with the practice criteria were judged inappropriate, and "this judgment was communicated directly to the physicians involved, either in conference or by the department chairman or the director of perinatology" (551).

Offering economic incentives to reduce Cesarean delivery rates (by equalizing payments for Ce - sarean delivery and vaginal birth), in contrast to the above interventions. has not been successful in the few studies that have been done (422). The success of noneconomic interventions and failure of economic explanations for rates of Cesarean delivery reinforce the idea that a variety of forces are important in determining patterns of clinical practice.

### *Anesthesia Monitoring Guidelines*

The development and dissemination of standards for patient monitoring during anesthesia have been credited with influencing clinical practice, reducing anesthesia-related deaths. and reducing malpractice insurance premiums (21 1,425.608). In 1987, a major malpractice insurance carrier in

[5] Profiling and feedback may hav c failed because no individual-level data were provided: the involved physicians chose to decline this information ( 470).

Massachusetts offered a 20-percent reduction in premiums for physicians who agreed in writing to follow guidelines, developed by the Risk Management Foundation of Harvard Medical Institutions, intended to reduce anesthesia-related deaths. The insurer reported a reduction in anesthesia-related claims following implementation of the program (608). After the guidelines were subsequently adopted as national standards by the American Society of Anesthesiologists, malpractice insurers across the country reported similar results (608).

Not all of the reductions in the number of anesthesia-related claims can be attributed to the guidelines. Over the past decade, a general downward trend in rates of adverse anesthesia events has been observed, possibly related to the guidelines but possibly also related to other changes in the availability and use of new technologies (580,608,664). A number of other factors may have contributed to the apparent success of the anesthesia monitoring standards as well:

- The guidelines were purposefully promulgated as minimal criteria for adequate practice (212).
- The standards describe the recommended practice in simple and explicit terms.
- The practices are simple to undertake and require inexpensive. readily available equipment.
- The burden of adherence is trivial compared with the potential catastrophic consequences of failure to comply.
- Economic incentives exist to follow the guidelines (i.e., reduced malpractice insurance premiums).

### *Hospital Length-of-Stay Guidelines*

Decisions about the duration of time that a patient stays in the hospital, or in a special care unit, involve a different set of considerations about risks and benefits than decisions about whether to order a particular drug or test. Thus, guidelines about length of stay maybe affected by different factors as well.

In one study investigating the impact of, guidelines on the length of stay in the intensive care unit (ICU) (193), guidelines were developed internally by staff with the input of respected outside experts. These guidelines were part of daily discussions on the care for each ICU patient. In addition, a weekly review of patient care was conducted, with focused exploration of all cases in which the guidelines were not followed. The intervention reduced the ICU length of stay in one subset of patients (those who were admitted to the ICU primarily for observation), but not others. The study suggested that guidelines on length of stay, developed locally and applied intensively in a teaching setting, can influence how long certain types of patients are kept in the ICU.

Another study examined the impact of a guideline promoting early discharge for low-risk patients admitted to the hospital with chest pain. In this study, patients identified by the guidelines as being "low risk" had a note placed on their chart indicating that they could be discharged within 48 hours, with little risk to their health, if their tests did not confirm a heart attack (902,905). If the patient remained more than 24 hours despite the note, telephone contact was made by a utilization management physician (a respected local practitioner) who reinforced the guideline recommendation. This intervention resulted in a reduced length of stay and estimated savings of $1,400 per low-risk patient, without any detected adverse effects on a wide range of measured patient outcomes (902,905). However, the impact on length of stay did not persist when feedback was discontinued, suggesting that the note and phone call served as necessary reminders to follow the guideline (902,905).

Interestingly, an identical intervention for patients with congestive heart failure (CHF) in the same hospital was not successful in reducing the length of stay (901). Several differences between the interventions may explain the differences in study outcomes. First, the guideline for CHF patients was more complex and less easily understood than the chest pain guideline. Second, the CHF guideline was developed by a local investigator using local data, but was not shown in independently published studies to be an accurate tool

for safely managing the site of care of heart failure patients. In contrast. the chest pain protocol was based on studies published in the pm--reviewed literature.

### *Drug Prescribing and Test Ordering*

Use of pharmaceuticals and diagnostic tests arc areas where changes in practice may have considerable potential for reducing the costs of care, but experience in using guidelines to affect practice in this area are mixed. For example, physician practices have been shown not to correlate well with recommended use of antibiotics to prevent heart valve infections in patients undergoing procedures that might introduce bacteria into the blood (87,634). Dissemination of guidelines on antibiotic selection in another study failed to influence use of these antibiotics (671 ). One study did report a decrease in unnecessary use of albumin therapy for hypovolemia in France with repeated distribution of a guidelinc (189). Guidelines have also been used successfully to reduce the use of x-rays (262,66 1 ) and diagnostic blood tests (673, 718).

Guidelines for use of blood products have been a high priority interest area, Transfusion practices of surgeons were altered through a detailing intervent ion based on guidelines from a national blood banking organization (7 16). However, in some cases, adverse outcomes may dissuade physicians from following new guidelines. For example, a hospital-based intervention intended to reduce use of platelet transfusions caused two episodes of bleeding in patients with low platelet counts, which eliminated physician compliance with the guideline (695).

### *Clinical Management Guidelines*

Many of the guidelines noted above target specific components of patient care or specific interventions. giving relatively explicit recommendations for the use of selected services. other guidelines, however, focus mow broadly on management of a giving clinical condition over time. Such guidelines tend to be more complex and less explicit than procedure-oriented guidelines.

Chart review studies suggest that, in many cases, current clinical practices do not conform to the specifications of existing clinical management guidelines. In treatment of pediatric diarrhea. for example, a guideline recommendation for rapid dehydration was followed properly by only a minority of pediatricians (60). Another study on guidelines for cancer staging and radiation therapy consultation showed a low rate of compliance despite the fact that the physicians who participated in the study also developed the guidelines (261 ).

Greater compliance was achieved in a program that used preprinted order sheets based on guidelines for management of pediatric asthma (93). The order sheet specified nursing ratios, oxygen monitoring, drug dosing schedules. and laboratory orders. Pediatricians in the protocol group had changes in practice consistent with the protocol. Guidelines that can be reduced to explicit algorithms that outline care for a complex and serious medical problem were in this case a well accepted intervention, although the study was not able to show whether compliance with the guideline improved patient outcomes (818).

A guideline-based intervention for management of patients with peptic ulcer disease that used a variety of administrative strategies was successful in altering drug prescribing practices and reducing the use of' endoscopy (93). Physicians were encouraged to use a less expensive ulcer medication by requiring prior authorization for the more expensive alternative. Fewer endoscopies were performed after active implementation of locally modified guidelines from the American Gastroenterological Association (93).

## Factors Influencing the Impact of Guidelines

Because of major limitations in the literature, it is difficult to draw firm conclusions as to which factors are the most influential in promoting practice change via guidelines. First, studies on the impact of guidelines are often poorly documented, and in many cases they arc inaccurately characterized in review articles. Second, because many of the in-

terventions studied involve multiple variables (e.g., an intervention may use both educational and financial incentives), it is difficult to isolate which of the features were important and which were irrelevant to the outcomes. Only on rare occasions are comparisons made between interventions that differ according to only one variable. Third, much of the "evidence" is in the form of anecdotal comparisons between guidelines implementation and broader utilization trends. Such comparisons do not control for other factors that may have influenced changes in utilization rates and are therefore not compelling evidence of the success of guidelines.

Taken collectively, however, the guideline implementation activities reviewed above and by others suggest that a variety of factors are important to successful implementation. Many of these factors mirror the elements of the influences on physician behavior discussed earlier in this chapter. They include:

- **Intensity of dissemination efforts-Local dissemination efforts (e.g., computerized protocols, formulary restrictions, telephone reminders) increase the likelihood that national guidelines will be followed (153,298). Multiple strategies may be needed to bring about a desired practice change. To be successful, implementation models must be tailored to address the variety of influences relevant to a given practice.**
- **Follow-through of dissemination efforts—Continued efforts may be necessary to ensure that guidelines have a lasting impact on clinical practice. It is well recognized that even when changes in practice occur, they often do not persist beyond the period of intervention (216,679,760).**
- **Type of clinical problem and clinical tasks addressed by the guideline—Guidelines that attempt to alter a single, well defined practice (e.g., performance of a screening test) seem to be more effective in changing practice than more complicated or general guidelines addressing more complex behaviors. Physicians may also be more compelled to follow guidelines that aim to prevent potentially catastrophic outcomes through the use of relatively simple and inexpensive interventions (as in the anesthesia example). Guidelines that promise only marginal improvements in outcomes or that prescribe new practices that are difficult or expensive to implement maybe less successful at changing practice.**
- Source of the guideline—Guidelines issued by respected sources seem to be more easily accepted (768). Physicians are most likely to distrust guidelines developed by organizations that they perceive to be biased or to have a financial interest in the recommendations.
- Physician participation in the development or adoption of the guideline—Physician involvement in the development or adoption of guidelines used in the intervention seems to increase the likelihood of successful implementation. Even simply being included in the deliberation process enhances the clinicians' sense of participation.
- **Form and specificity of the guideline's recommendations--Guidelines that are simple, clear, and explicit are more likely to be effective than ones that are complex and leave much room for interpretation or judgment (307).**
- **Legal considerations—Guidelines that suggest a minimum standard of care maybe particularly influential, because physicians may believe they suggest a legal standard.**
- **Financial and administrative mechanisms—These can be powerful tools to reinforce guidelines but if used alone often have unwanted side effects.**
- **Strength of evidence on which the guideline is based—Physicians maybe influenced by the strength of evidence on which a guideline is based. Also, if a protocol has proven effective at a number of different sites, cl inicians may be more likely to adopt its prescriptions. Guidelines for screening tests backed up by strong, consistent evidence, for example, may be more influential than those based on weaker or more controversial evidence.**

There are a number of instances where high-quality evidence from clinical trials appears to have contributed to the success of guideline implementation. For example, in the case of pacemakers, consistent, strong, and high-quality evidence supported a clear and explicit recommendation that was reinforced by payment policy (304). Similarly. efforts to reduce Cesarean delivery rates in Ontario and at Mt. Sinai Hospital in Chicago (470,55 1 ) were buttressed by strong studies showing that outcomes for mother and child were not worse when vaginal deliveries were attempted (250). In contrast, the lack of good quality and widely accepted evidence supporting guideline recommendations might explain some of the failures to implement the guideline to reduce critical care unit stays for patients with heart failure (904).

Guidelines can sometimes be successfully implemented even if based on studies with less-than-ideal designs, however, if the likely effect from changing practice is sufficiently large and the outcome to be averted a severe one. In contrast to the guidelines on Cesarean delivery. which were supported by a number of high-quality studies, the anesthesia guideline recommendations were based primarily on a review of a case series of adverse events (21 2).

The relationship between quality of evidence and impact on practice has not been well studied. However. as discussed in chapter 7, consistent high-quality evidence improves expert group consensus, which in turn enables the panels of experts to make strong, clear recommendations about appropriate practice (470). It also presumably lessens the chance that the recommendations of separately developed guidelines will conflict and confuse users.

## CONCLUSIONS

Physicians' (and patients') decisions are shaped not only by information on the potential risks and benefits of possible alternative therapies, but also by a range of external forces that include economic and legal pressures, the wishes of patients and their families, the practice styles and input of colleagues and local experts, and the set of administrative rules within which physicians operate. Because of the multiple forces affecting clinical decisions, no single strategy to change physician practice is likely to be identified. For any given clinical situation. a few sources of influence on practice are likely to predominate. Efforts to alter practices may be most efficiently achieved when the forces that shape those practices are identified, and this knowledge is used to select and design an intervention to alter that practice.

Financial and administrative mechanisms can have a potent influence on physician and patient behavior, but they are often unselective and can have unintended, undesired secondary effects. As a means of implementing guidelines they do provide motivation, but their benefit depends on the degree to which secondary impacts are minimized and on the validity of the guidelines themselves.

Patients are involved in clinical decisions, and patient demand influences physician behavior. Educational strategies aimed at patients can be successful in altering clinical practices to improve the benefit to the patient, though health care costs do not necessarily decrease as a result.

Physician decisions are strongly affected by the beliefs and practices of their clinical colleagues. Physicians are more likely to receive and incorporate information from sources they know and respect, or from organizations with which they are associated. Informal networks of information through local medical communities have been shown to be important routes of information exchange among clinical providers. In contrast. physicians typically doubt or reject information that comes from distant sources, particularly if they have any reason to suspect the presence of a bias they do not share.

Physician involvement in the process of change appears to improve the success of interventions to alter practices. This has been observed in small group educational workshops, practice auditing with feedback. clinical guideline development, conduct of clinical trials. and continuous quality improvement.

Data collected in the course of patient care are being used. often effectively, to change practice via practice profiling, feedback, and CQI. However, none of these applications can provide reliable evidence regarding which alternative medical practices work best unless they are combined with evidence (alone or interpreted through guidelines) on the most effective practices.

Although even passive dissemination of information often plays an enabling role for future efforts, the development and distribution of guideline statements is often ineffective in altering physician practices, no matter how important the issue or how valid the guideline method. Specific strategies to implement guideline recommendations are necessary to ensure practice change, and the existing studies suggest that more intensive efforts to alter practice are generally more successful. Implementation of guidelines will require considerable effort and skill. Implementation is also more effective if tailored to the particular condition, type of clinical task, and patient population involved.

Impact on practice change is increased by strong, well defined recommendations for a specific practice. Firm recommendations are enhanced by strong and consistent evidence. Thus, where there is variation in clinical practices and no good evidence about which practices are most effective, performing high-quality clinical trials is important not only to resolve the question but to convince physicians of the answer.

The literature on physician behavior suggests that locally developed guidelines may be more influential than guidelines issued from remote sources. However, many physicians may prefer a flexible set of national guidelines to a panoply of more specific and potentially conflicting guidelines from payers, local government, institutions, and other entities. The challenge for federal guidelines efforts will be finding effective means of disseminating their guidelines through the local networks of communication that have been so effective in influencing physician practice. Allowing for some local adaptation and modification of clinical strategies developed outside physician communities would help ensure that the information has an effect on practice.

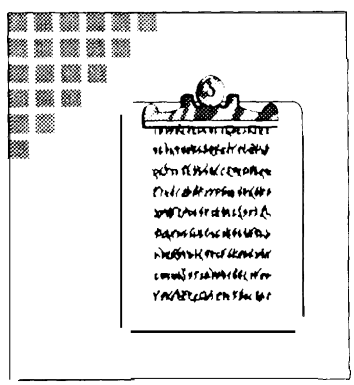
Additional research is needed on the forces and strategies that influence clinical decisionmaking. It would be particularly useful to have studies that carefully control all but a single aspect of an intervention, in order to begin to identify specifically the elements required for effective dissemination. It is also important to test strategies for categories of clinical practice other than the few that have been extensively studied to date and to apply practice change interventions to physician populations other than residents and medical students. Evaluations of practice change interventions should carefully document all of the forces and factors associated with the intervention so that its impact can be interpreted with full knowledge of all variables that might account for observed effects.

# Appendix A: How This Study Was Conducted

A

The Office of Technology Assessment (OTA) was first asked to examine issues relating to effectiveness research and health technology assessment in a July 1992 letter from Senators Kennedy and Hatch, on behalf of the Senate Committee on Labor and Human Resources. That letter asked that OTA "conduct an evaluation of the field of health technology assessment, identify the strengths and weaknesses of current efforts, and outline options that 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 guidelines development, and others." Subsequently, Senator Grassley (of the congressional Technology Assessment Board) and Congressman Dingell (on behalf of the House Committee on Energy and Commerce) sent letters to OTA supporting this study (176,294).

OTA's congressional Technology Assessment Board approved the proposed OTA study of "Prospects for Health Technology Assessment" in August 1992. The study began two months later on October 1. Products associated with this study are this report *(Identifying Health Technologies That Work: Searching for Evidence)* and five background papers describing recent methodological advances and issues in greater detail (forthcoming).



## THE ADVISORY PANEL

OTA advisory panels provide advice to project staff regarding the scope and direction of the project, possible study approaches, important perspectives to consider, and important resources to consult. Panel members also review drafts of documents produced

during the course of the study. No attempt is made to develop consensus among panel members; in fact, a wide diversity of views is sought. Although panel members provide indispensable advice and assistance, they do not write the report. OTA retains full responsibility for the contents and conclusions of its publications.

The advisory panel for this study initially consisted of 15 members with expertise ranging from ethics, medicine, economics, and law to research methodologies, clinical evaluation, and policymaking. Perspectives represented on the panel included third-party payers; manufacturers; researchers: clinicians; and defenders of several different intellectual approaches to clinical evaluation. A list of panel members can be found at the front of this report. One member resigned from the panel part way through the study after taking a position with the executive branch. (According to OTA policy, employees of the federal government participate in studies as observers, reviewers, and workshop participants but do not serve on OTA advisory panels. )

The advisory panel met twice during the course of the study. On February 9, 1993, panel members discussed the scope and approach of the study. On October 25, 1993, the panel reviewed and discussed curly drafts of some sections of the report.

## INFORMATION SOURCES

The four types of information on which this study relied most heavily were the published health care literature; information provided by various federal agencies and offices about their activities; information from personal conversations, correspodence. presentations, and workshops; and contracted background papers on particular research methodologies.

### Published Literature

Relevant published literature was identified through computerized searches and from the suggestions of the many people consulted during the course of this study. Because the scope of the study was quite broad, no single formal literature search strategy was defined and used. Given the diversity of relevant topics to this assessment, any such search would have yielded an unmanageable number of articles. Instead, OTA limited its search strategy to computerized searches of the health literature on selected topics (e.g., commentaries and reviews of cost-effectiveness analysis) and acquisition of publications suggested to project staff or cited in other relevant publications. Staff also monitored newsletters and other periodicals on effectiveness research, guidelines, and other relevant topics. The expertise of panelists, other reviewers, and contractors helped ensure that important publications were not missed.

### Information on Federal Activities

OTA maintained ongoing contact with the Agency for Health Care Policy and Research and obtained from the agency information such as guidelines documents, lists of projects in the agency Medical Treatment Effectiveness Program, and other documents and information. OTA staff also met with the directors of many offices within the agency.

To obtain information on relevant activities elsewhere in the federal government, OTA sent letters to the administrators and directors of the Centers for Disease Control and Prevention, Health Care Financing Administration, each institute in the National Institutes of Health, and liaison staff at the Department of Veterans Affairs. The letters described the OTA study and asked for information on activities conducted or sponsored by those agencies that, in the opinion of that person, might be relevant to the study.[1] Similar information was requested of staff at the Office of Disease Prevention and Health Promotion by telephone. All agencies were responsive to the OTA request. with varying levels of detail.

[1] OTA staff contacted personnel in the Department of Defense as well but, due to resource constraints, did not pursue information on relevant activities of this department in any detail.

Information provided formally by the various agencies was augmented with staff-to-staff discussions, and published and unpublished information from other sources.

## Workshops, Conferences, and Personal Contacts

As part of this study, OTA held two workshops. The first of these. cosponsored by the OTA study of *Technology, lnsurunce, and the Health Care System,* was on the topic of "Alternative Visions for Using Effectiveness and Appropriateness Information to Design Health Benefits: Implications for Health Care Reform and Technology Assessment" and took place on January 26, 1993. The purpose of the workshop was to explore and discuss how information on clinical effectiveness might be used in designing insurance benefits under some of the proposals for health care reform then being put forth.

A second workshop, on "Linking Medical Evidence With Clinical Practice: Progress and Barriers," took place on May 18, 1993. The goal of this workshop was to bring together researchers and practitioners to discuss barriers to the practical implementation of clinical practice guidelines and other forms of evidence about medical care. Information obtained in preparation for and during this workshop became the basis for much of chapter 8 of the report. Workshops participants for both workshops are listed in appendix B.

OTA staff also attended, as observers, numerous workshops and conferences on relevant topics sponsored by other organizations during the course of this study. Examples include:

- "The Role of Outcomes Research in the Reformed Health Care System" (University of Maryland Center for Health Policy Research, Oct. 4-6, 1992)
- "Medical Effectiveness Research and Clinical Practice Guidelines: Implications for State Governments" (AHCPR, NOV. 9-11, 1992)
- "Medical Effectiveness Research: Strategies for the Future" (AHCPR, Feb. 17-18, 1993)
- "Doing More Good Than Harm: The Evaluation of Interventions" (New York Academy of Sciences, Mar. 22-25, 1993)
- "Health Services Research: Implications for Policy, Management, and Clinical Practice" (Association for Health Services Research, June 27-29. 1993)
- "Forum Levels of Evidence Workshop" (AHCPR, NOV. 1-2, 1993)
- "Cost Analysis Methodology for Clinical Practice Guidelines" (AHCPR, Nov. 22-23, 1993)
- '*An Evidenced-Based Health Care System: The Case for Clinical Trials Registries" (NIH, Office of Medical Applications of Research, Dec. 6-7, 1993)
- "Research Synthesis: Social Science Informing Public Policy" (Russell Sage Foundation, June 21• 1994)

In addition to the discussion and presentations at panel meetings, workshops. conferences, and other organized events, OTA staff contacted numerous individuals to discuss issues, evidence, and experiences related to the topics covered in this report. Staff also attended relevant congressional hearings.

## Contracted Papers

OTA commissioned with five individuals to provide detailed, policy relevant papers on five methodological topics of particular interest or importance to this study. The topics and contractors were:

- Measuring health status and quality of life through patients' reports: Floyd J. Fowler, University of Massachusetts, Boston, MA
- Analysis of large administrative and clinical databases: Jeffrey Whittle, Pittsburgh VA Hospital, Pittsburgh, PA
- Meta-analysis: Matthew P. Longnecker, University of California, Los Angeles, CA
- Large and simple clinical trials: Julie E. Buring, Michael A. Jonas. and Charles H. Hennekens, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

- Economic data and analysis in clinical trials: Neil R. Powe and Robert I. Griffiths, Johns Hopkins University, Baltimore, MD

These five background papers are being published separately in a forthcoming volume. OTA also contracted with Craig Tanio, University of Pennsylvania, to provide some analysis and background regarding the medical evidence used for some clinical practice guidelines.

## REVIEW PROCESS

An early draft of sections of the report was discussed at the advisory panel meeting in October 1993. The full draft was sent for review to advisory panel members in March 1994 and to approximately 100 outside experts for comment the following month. One-fourth of these experts were federal officials. The reminder comprised researchers, health insurers, technology assessment and benefits consultants, clinicians, and others with relevant expertise and perspectives.

A final draft, revised after considering all reviewer comments, was submitted to the Technology Assessment Board in June 1994.

# Appendix B: Acknowledgments

B

*OTA wishes to thank the advisory panel to the "Prospects for Health Technology Assessment" study (which led to this report), its contractors, and the individuals and organizations listed below for their assistance in the course of this study. These individuals and organizations do not necessarily approve, disapprove, or endorse this report. OTA assumes full responsibility for the report and for the accuracy of its content.*

**Participants in the Workshop on "Alternative Visions for Using Effectiveness and Appropriateness Information to Design Health Benefits: Implications for Health Care Reform and Technology Assessment"**

**Naomi Aronson**
Medical Necessity Program
Blue Cross and Blue Shield Association
Chicago, IL

**Robert Brook**
*The* RAND Corp.
Santa Monica, CA

**Stuart M. Butler**
The Heritage Foundation
Washington, DC

**Arthur Caplan**
Center for Biomedical Ethics
University of Minnesota
Minneapolis, MN

**Paul Ellwood**
Jackson Hole Group
Jackson. WY

**Beth Fuchs**
Congressional Research Service
U.S. Congress
Washington, DC

**Mark A. Hall**
Arizona State University
Tempe, AZ

**Lucian L. Leape**
Harvard School of Public Health
Boston. MA

**Patricia Nazemetz**
Xerox Corp.
Stamford. CT

**Paige Sipes-Metzler**
Oregon Health Services Commission
Portland, OR

**Gordon Trapnell**
Gordon Trapnell Actuaries
Annandale, VA

## Participants in the Workshop on "Linking Medical Evidence with Clinical Practice"

**John Eisenberg**
***Chairperson***
Department of Medicine
Georgetown Medical Center
Washington, DC

**Thomas Backer**
Human Interaction Research Institute
Los Angeles, CA

**Catherine Borbas**
Healthcare Education and Research Foundation
St. Paul, MN

**Elizabeth Cobbs**
George Washington University
Washington, DC

**Arnold Epstein**
The White House
Washington, DC

**Alice Gosfield**
Alice Gosfield and Associates
Philadelphia, PA

**John Harris**
CaliforniaCare
Woodland Hills, CA

**Alan Hillman**
Leonard Davis Institute of Health Economics
University of Pennsylvania
Philadelphia, PA

**Sherrie Kaplan**
The Health Institute
New England Medical Center
Boston, MA

**Robert Keller**
Maine Medical Assessment Foundation
Augusta, ME

**Risa Lavizzo-Mourey**
School of Medicine
University of Pennsylvania
Philadelphia, PA

**Charles Silberman**
New York, NY

**Stephen Soumerai**
Harvard School of Medicine
Boston, MA

**Michael Stuart**
Group Health of Puget Sound
Seattle, WA

**Scott Weingarten**
Cedars-Sinai Medical Center
Los Angeles, CA

## Respondents to OTA's Request for Information on Relevant Federal Activities

**Agency for Health Care Policy and Research**
Rockville, MD
Jill Bernstein

**Centers for Disease Control and Prevention**
Atlanta, GA
Steven Teutsch

Health Care Financing Administration
Washington, DC
Bruce Vladeck

**National Institutes of Health**
Bethesda/Rockville, MD:

- *Clinical Center*
  Saul Rosen
- *Division of Research Grants*
  Jerome Green
- *Fogarty International Center*
  Philip Schambra
- *National Cancer Institute*
  Daniel Idhe
- *National Center for Human Genome Research*
  Elke Jordan
- *National Center for Nursing Research*
  Ada Sue Hinshaw
- *National Center for Research Resources*
  Judith Vaitukaitis
- *National Eye Institute*
  Carl Kupfer
- *National Heart, Lung, and Blood Institute*
  Claude L'Enfant

- *National Institute on Alcohol Abuse and Alcoholism*
  Enoch Gordis
- *National Institute of Allergy and Infectious Diseases*
  Anthony Fauci
- *National Institute of Arthritis and Musculoskeletal and Skin Diseases*
  Steven Hausman
- *National Institute on Child Health and Human Development*
  Duane Alexander
- *National Institute on Deafness and Other Communication Disorders*
  James Snow
- *National Institute of Dental Research*
  Harald Loe
- *National Institute of Diabetes and Digestive and Kidney Diseases*
  Richard Farishian
- *National Institute on Drug Abuse*
  Richard Mill stein
- *National Institute of Environmental Health Sciences*
  Jerry Phelps
- *National Institute of General Medical Sciences*
  Ruth Kirschstein
- *National Institute of Mental Health*
  Frederick Goodwin
- *National Institute of Neurological Disorders and Stroke*
  Murray Goldstein

### Other Federal Staff

**Agency for Health Care Policy and Research**
Rockville, MD
Irma Arispe
Emilie Baebel
Carolyn Clancy
Cathy Crawford
Michael Fitzmaurice
Barbara Fleming
Richard Greene
Carole Hudgings
Claire Maklan
Kathleen McCormick
Eleanor Perfetto
Margaret van Amarange

**Centers for Disease Control and Prevention**
Atlanta, GA
Dixie Snider

**Food and Drug Administration**
Rockville, MD
Lois Adams
Jim Bilstad
John Harter
Jim Hooten
Lou Morris
James G, Norman
Robert O'Neill
Kate Sheridan
Joy Stevens
Robert Temple
Dave West
Ann Witt
Randy Wykoff

**Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion**
Washington, DC
James Harrell

**Department of Veterans Affairs Veterans Health Administration**
Washington, DC
Shirley Meehan

**U.S. Congress Congressional Research Service**
Washington, DC
Bernice Reyes-Akinbileje

**U.S. General Accounting Office**
Washington, DC
George Silberman

**Health Care Financing Administration**
Baltimore, MD
Jennifer Boulanger

**Office of Disease Prevention and Health Promotion**
Washington, DC
Marthe Gold
Joanna Siegel
Steven Woolf

**National Institutes of Health**
Bethesda, MD
- *National Cancer Institute*
  Nancy Breen
  Martin Brown
  Betsy Duane
  Mitchell Gail
  Michael Friedman
  Susan Hubbard
  Larry Kessler
  Barnett Kramer
  Susan Nay field
  Richard Simon
  Edward Sondick
  J. Paul Van Nevel
- *National Center for Human Genome Research*
  Mary Lawrence
- *National Heart, Lung, and Blood Institute*
  .Ccarol Roth
  Gregory Morosco
- *National Institute on Aging*
  Barbara Kellner
- *National Institute of Allergies and Infectious Diseases*
  Linda Rosendorf
- *National Institute of Child Health and Human Development*
  Monica Rowe
- *National Institute of Diabetes and Digestive and Kidney Diseases*
  Dee LeRoy
- *National Institute of Environmental Health Sciences*
  Jerry Phelps
- *National Institute of Mental Health*
  Grayson Norquist
- Office *of the Director*
  William Harlan
  Anne Houser
  Kristen Kizer
  Jay Moskowitz
  James Schuttinga

**Department of Defense**
Office of the Air Force Surgeon General
Washington, DC
Pricilla Worrel

**Department of Veterans Affairs**
Washington, DC
Daniel Deykin

## Other Acknowledgments

**Michael Barry**
Massachusetts General Hospital
Boston, MA

**Marion Becker**
Center for Health Systems, Research and Analysis
University of Wisconsin
Madison, WI

**Larry Bierley**
Anne Arundel Medical Center
Annapolis, MD

**Joel Bobula**
Pharmaceutical Research and Manufacturers of America
Washington, DC

**Mike Boushon**
Bureau of Health Care Financing
Madison, WI

**Philip Boyle**
Hastings Center
Briarcliff Manor, NY

**Irene Brandt**
Eli Lilly and Co.
Indianapolis, IN

**Vivian Coates**
ECRI
Plymouth Meeting, PA

**Abraham Cockett**
American Urological Association
Baltimore, MD

**Stephen Coons**
College of Pharmacy
University of Arizona
Tucson, AZ

**Dennis Cotter**
Medical Technology and Practice Patterns Institute, Inc.
Washington, DC

**John Cova**
Health Insurance Association of America
Washington, DC

**Dave Davis**
Faculty of Medicine
University of Toronto

**Kay Dickerson**
School of Medicine
University of Maryland
Baltimore, MD

**Ralph DiDomenico**
Pfizer Hospital Products Group
New York, NY

**Dan Dragalin**
Mendham, NJ

**Mary Fahey**
Department of Health
Minneapolis, MN

**Mark Fenderick**
University of Michigan
Medical Center
Ann Arbo, MI

**Marilyn Field**
Institute of Medicine
Washington, DC

**Elliot Fisher**
Dartmouth Medical School
Hanover, NH

**Alan Garber**
National Bureau of Economic Research, Inc.
Stanford, CA

**Susan Gleeson**
Blue Cross and Blue Shield Association
Chicago, IL

**William Golden**
University Hospital of Arkansas
Little Rock, AK

**Cliff Goodman**
Washington, DC

**Larry Gottleib**
Harvard Community Health Plan
Brookline, MA

**Gordon Guyatt**
McMaster University
Hamilton, Ontario

**David Hadorn**
RAND Corp.
Santa Monica, CA

**Mark Hall**
School of Law
Wake Forest University
Winston-Salem. NC

**David Haugen**
Department of Health
Minneapolis, MN

**Brian Haynes**
Health Information Research Unit
McMaster University
Hamilton, Ontario

**Hy Kahn**
U.S. Healthcare
Philadelphia, PA

**Lorrie Kalasinski**
Indiana University
Indianapolis, IN

**Sona Kalousdian**
Department of Technology Assessment
American Medical Association
Chicago, IL

**David Kanouse**
RAND Corp.
Santa Monica, CA

**Alison Keith**
Pfizer, Inc.
NewYork,NY

**John Kelly**
GMIS
Malvem,PA

**Peter Kongstvedt**
Blue Cross and Blue Shield of the National Capitol Area
Washington, DC

**Jacqueline Kosecoff**
Value Health Sciences Inc.
Santa Monica, CA

**Louis Lasagna**
School of Medicine
Tufts University
Medford, MA

**Roz Lasker**
Washington, DC

**Jeffrey Lerner**
ECRI
Plymouth Meeting, PA

**Ian Leverton**
Kaiser Permanente
Oakland, CA

**Alessandro Liberati**
Instituto Ricerche
Farmacologiche "Mario Negri"
Milan, Italy

**Candace Littell**
Health Care Technology Institute
Alexandria, VA

**Kathleen Lohr**
Institute of Medicine
Washington, DC

**Bryan Luce**
Battelle Institute
Arlington, VA

**Karl Matuszewski**
University Hospital Consortium
Oak Brook, IL

**William McGivney**
Aetna Health Plans
Hartford, CT

**Morris Melloy**
National Association of
Insurance Commissioners
Kansas City, MO

**Stephanie Mensh**
American Urological Association
Baltimore, MD

**Pamela Middlestat**
Group Health Association of
America
Washington, DC

**Michael Millenson**
Center for Health Services Policy
and Research
Northwestern University
Evanston, IL

**Brian Mittman**
RAND Corp.
Santa Monica, CA

**Curt Mueller**
Project HOPE
Chevy Chase, MD

**Cynthia Mulrow**
Audie L. Murphy Memorial
Veterans Hospital
San Antonio. TX

**David Nash**
Thomas Jefferson University
Hospital
Philadelphia, PA

**Michael Neff**
MediCal Program
Sacramento, CA

**Dorit Opher**
ECRI
Plymouth Meeting, PA

**Tracy Paiva**
Kaiser Permanente of Northern
California
Oakland, CA

**William Peters**
Duke University
Durham, NC

**Gene Reeder**
University of South Carolina
Columbia, SC

**Henry W. Reiken**
Kezar Associates
Washington, DC

**William L. Roper**
The Prudential Center for Health
Care Research
Atlanta, GA

**Joan Rovira**
SOIKAS
Barcelona, Spain

**Gordon Schatz**
Reed, Smith, Shaw, & McClay
Washington, DC

**Steven Schoenbaum**
Harvard Community Health Plan
Brookline, MA

**David Schriger**
UCLA Emergency Medicine Center
Los Angeles, CA

**Paul Shekelle**
RAND Corp.
Santa Monica, CA

**Jane Sisk**
School of Public Health
Columbia University
New York, NY

**Harold Sox**
Dartmouth Medical School
Lebanon, NH

**Earl Steinberg**
Johns Hopkins University Medical School
Baltimore. MD

**Chester Strobel**
Pfizer, Inc
Groton, CT

**Albert Siu**
New York State Department of Health
Albany, NY

**Sean Sullivan**
College of Pharmacy
University of Washington
Seattle, WA

**Leigh Thompson**
Eli Lilly and Co.
Indianapolis, IN

**Ralph Turner**
Pfizer. Inc.
Groton. CT

**Carol Vervoncoer**
Technology Assessment Group, Inc.
San Francisco, CA

**Karen Watson**
Kaiser Permanence of Northern California
Oakland, CA

**Linda White**
American College of Physicians
Philadelphia, PA

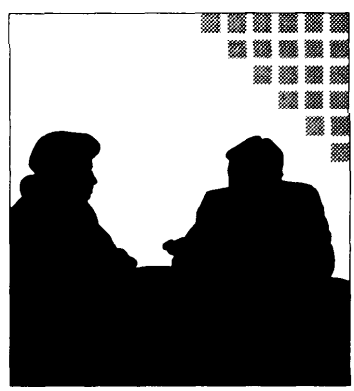
# C Appendix C: Guideline Development Activities

This appendix describes how six core tasks are accomplished by several of the principal federal and private organizations sponsoring the development of guidelines:

1. selecting a guideline topic,
2. defining the purpose and scope of the guideline,
3. collecting and synthesizing evidence,
4. devising a method to deliberate and then make judgments and recommendations,
5. reviewing the guideline, and
6. updating the guideline.

The focus of this appendix is on major federal guidelines activities. Selected, well established private guideline efforts are also described, to put federal activities in a broader context.

## FEDERAL GUIDELINE ACTIVITIES



At the federal level, most clinical practice guidelines intended for general use are sponsored by agencies within the Department of Health and Human Services (DHHS), especially the Agency for Health Care Policy and Research (AHCPR) and the National Institutes of Health (NIH) (376). Within NIH, the most prominent guideline-like activities are the Consensus Development Conference Program, administered through the Office of Medical Applications of Research (OMAR); [1] guidelines issued through the National Education Programs of the National Heart, Lung, and Blood Institute (NHLBI); and the cancer information statements of the National Cancer Institute. The DHHS Office of the Assistant Secretary for Health, through its program on health pro-

[1] OMAR is administratively housed within NIH's Office of the Director.

**TABLE C-1: Guidelines Developed or Sponsored by AHCPR, 1992-July 1994**

| Topic | Initiation date | Release date |
|---|---|---|
| Acute pain management | July 1990 | February 1992 |
| Anxiety and panic disorder | November 1992 | in progress |
| Benign prostatic hyperplasia | July 1990 | February 1994 |
| Cancer-related pain | September 1991 | March 1994 |
| Cardiac rehabilitation[1] | May 1992 | in progress |
| Cataracts in adults | August 1990 | February 1993 |
| Colorectal cancer screening[1] | May 1994 | in progress |
| Congestive heart failure[1] | January 1992 | June 1994 |
| Depression in primary care | September 1990 | April 1993 |
| Early HIV infection | July 1991 | January 1994 |
| Lower back problems | November 1991 | in progress |
| Otitis media in children[1] | November 1991 | July 1994 |
| Post stroke rehabilitation[1] | January 1992 | in progress |
| Prevention of pressure ulcers | August 1990 | May 1992 |
| Quality determinants of mammography | June 1991 | in progress |
| Screening for Alzheimer's and related dementias | March 1992 | in progress |
| Sickle cell disease in infants | November 1990 | April 1993 |
| Smoking prevention and cessation | December 1992 | in progress |
| Treatment of pressure ulcers | June 1991 | in progress |
| Unstable angina[1] | May 1992 | in progress |
| Urinary Incontinence in adults | August 1990 | March 1992 |

[1]Guidelines were produced or are being produced by an AHCPR contractor

SOURCES 'Guidelines Being Developed, Agency for Health Care Policy and Research, Pubic Health Service U S Department of Health and Human Services Rockville MD, unpublished document, September 1993 E McGovern Agency for Health Care Policy and Research, Public Health Serv ce U S Department of Health Human Services Rockville MD personal communication Mar 4 1994 Physician Payment Review Commission *Annual Report to Congress, 1992* (Washington DC U S Government Printing Off Ice 1992)

motion and disease prevention, and the Centers for Disease Control and Prevention (CDC) also have a significant role in guidelines development.

## Agency for Health Care Policy and Research

The focus of federal clinical practice guideline development resides in AHCPR's Office of the Forum for Quality and Effectiveness in Health Care. Since its inception in 1989. AHCPR has published 11 clinical guidelines, with 10 more are under development as of July 1994 (table C-l). The purpose of AHCPR-sponsored guidelines is "to enhance the quality, appropriateness, and effectiveness of health care" (812).

AHCPR sponsors the development of clinical guidelines, rather than developing them in-house. AHCPR can appoint guideline panels, contract with groups to develop guidelines, or recognize guidelines developed by other organizations. Most panels to date have been appointed by AHCPR, but six guidelines have been or are being developed under contract (i.e., guidelines on otitis media, congestive heart failure, post stroke rehabilitation, cardiac rehabilitation, unstable angina, and colorectal cancer screen in:). Medical review

criteria are being developed based on three of AHCPR's guidelines (those on acute pain management, urinary incontinence in adults, and benign prostatic hyperplasia) (813).

### Topic Selection

Most AHCPR guidelines have focused on the diagnosis and management of clinical conditions (e.g., pressure ulcers, depression) rather than on the use of individual technologies or treatments. According to the agency, it considers six factors when selecting guideline topics (8 12):

1. potential for reducing clinically significant and unexplained variations in services and procedures used in the prevention, diagnosis, management, or outcomes related to the clinical condition;
2. number of individuals affected by the condition;
3. adequacy of scientific evidence with which to develop a guideline;
4. amenability of a particular condition to prevention;
5. specific needs of the Medicare and Medicaid populations; and
6. cost of the condition to all payers, including patients.

When reauthorized in 1992, AHCPR was also directed to consider evidence of inappropriate utilization of health care resources such as variation in the frequency or the kind of treatment provided (Public Law 102- 410).

AHCPR solicits opinions regarding possible topics of guidelines through Federal Register notices and guideline-related publications (79). Since 1992, AHCPR has held meetings to discuss potential topics with experts and representatives of a variety of groups and interests (53). AHCPR also reports that it is conducting a study to determine optimal methods for selecting guidelines (855).

### *Scope of Guidelines*

The scope of AHCPR-sponsored guidelines was defined only very generally in the initial legislation establishing the agency. Guidelines were to consist of a synthesis of the available literature, considering the comparative effects of alternative services on health and functional capacity (Public Law 101 -239). When Congress reauthorized the agency in 1992, it specifically encouraged the development of practice guidelines that would allow providers and patients to compare costs as well as benefits of alternative medical strategies (Public Law 102-410). Many of the guidelines to date (e.g., the guideline on urinary incontinence) include some kind of estimate of the cost of implementing the guidelines, but none have included formal cost-effectiveness analyses.

Some guidelines have addressed selected health system constraints that might affect guideline implementation. The sickle cell disease guideline, for example, discusses the limited number of counselors available to provide genetic counseling, and the HIV* guideline discusses lack of insurance coverage as a barrier to care (810,818).

AHCPR publishes provider and consumer versions of each guideline.

### *Collecting and Synthesizing Evidence*

AHCPR-sponsored guidelines have emphasized exhaustive, systematic literature searches. For the HIV guideline panel, for example, the National Library of Medicine searched 30 databases and retrieved 36,000 citations. Even the smallest bibliography the NLM has prepared to date (for the acute pain management guideline) included 5,500 articles. Ultimately, very few articles are sufficiently relevant to the guideline discussion to be considered seriously. For the urinary incontinence guideline, for example, only about 2 percent of articles were actually used (617).

[2]HIV is the human immunodeficiency virus, which causes AIDS.

Explicit criteria are used to identify and synthesize the literature. The cataract guideline, for example, describes how topics to be searched were selected, how databases were searched, what search strategies were used, what dates bounded the search, what limitations were imposed on the search (e. g., English-language only), and how unpublished and recent literature were retrieved (806).

Guidelines panels have differed in how they have extracted evidence from the literature. Some have used informal methods, assigning articles to panel members for review and synthesis, while others have used explicit criteria to initially screen articles, and then have standardized methodologic reviews. Panels use tables (called "evidence tables") to summarize important aspects of the literature (e.g., the research design used in each study), and some panels (e.g., the acute pain management panel) have explicitly rated the quality of evidence. Most panels describe the strength of evidence used in support of each guideline recommendation and conclusion, but designations of strength of evidence have varied from panel to panel (815). AHCPR has sponsored research and workshops to develop and promote better methods to synthesize and apply evidence to guidelines work (815).

At least one public forum is held by each guideline panel, usually early in the guideline development process in conjunction with the second panel meeting.

### *Characteristics of Group Members and Processes*

AHCPR-sponsored guideline panels have been multidisciplinary and have ranged in size from 12 to 18 members. Most panel members are health care providers, but their backgrounds vary. Physicians predominate on most panels,[3] but a number of nonphysician providers are usually also included (e.g., nurses, social workers, psychologists). Each panel has included at least one family practice physician, one nurse, and a "consumer representative. "

Experts in nonclinical fields have only rarely been included on panels. Only one panel, for example, has included an economist (the otitis media panel). Methodologists have generally served as advisors to the group rather than being a part of the panel.

Panel members meet about four times throughout the course of guideline development. During their meetings, explicit. structured group processes are not used and consensus, when achieved, has not been formally defined.

Some more structured processes have been used during some aspects of the guideline devel - opment process. For example, panelists used a formal process in selecting topics to be considered in the HIV guideline. The panelists compiled a master list of relevant topics, categorized them, and rated the topics according to six attributes (e.g., importance of the issue to consumers). Methodologists rated each topic in terms of the technical feasibility of addressing it with structured analytic approaches such as decision analysis or meta-analysis. The panelists' final selection of topics were guided by the rankings determined by these two sets of scores (818). Guideline recommendations are often illustrated using an algorithm that shows recommended steps in clinical management in a flow chart (see chapter 7, box 7-1 ).

Over the past four years, as experience has accumulated, the process has become more explicit. For example, a methodology manual is being developed that is intended to guide the process of future panels.[4]

[3]Physicians were actually in the minority for two of the first AHCPR-sponsored guidelines (pressure ulcers in adults and acute pain management). Both specialty and primary care providers have been represented on each panel.

[4]AHCPR staff state that an early methodological guideline is outdated and no longer describes the process accurately (50 I ]. A new methodology manual has apparently been under development since 1992 but is not yet available.

AHCPR has several activities in progress or planned that will support its guidelines activities (813):

- An analytic unit within AHCPR will analyze cost data and baseline information on service utilization.
- AHCPR has commissioned a study on sources of cost data for guideline development and an evaluation of cost analyses conducted for 10 of the AHCPR-sponsored guidelines.

### *Review Process*

Draft guidelines are usually reviewed at least twice by outside experts and representatives of professional and consumer organizations. Reviewers are asked to assess the validity, reliability, clarity, clinical applicability, and utility of the guideline. Consumer brochures are reviewed by consumer representatives. For the HIV guideline, for example, HIV-positive individuals and parents or guardians of HIV-positive children participated in a focus group to evaluate the consumer guides.

Guidelines are also "pilot" tested by clinicians in practice settings. Here the guidelines are evaluated in terms of their clarity, clinical applicability, flexibility, resources and training needed to implement the guideline, and cost implications.

### *Development Time and Cost*

AHCPR guidelines have taken from one and a half to three and a half years to develop and have cost from $0.5 to $1 million. (This cost estimate excludes AHCPR staff-associated costs, and publication and dissemination costs (252).)

### *Updating Guidelines*

Most AHCPR guidelines have not specified update timetables, but at least one guideline (on cataracts) has included a scheduled review date (i.e., that the guideline should be revised in two years).[5]

## NIH Consensus Development Conference Program

NIH's OMAR has issued nearly 100 consensus statements since its inception in 1977, with 21 of them held since 1990 (table C-2). The primary mission of the Consensus Development Conference Program is to identify clinically relevant findings emerging from NIH research and to disseminate these findings to clinicians (237). Some consensus statements are limited in scope, but many are quite comprehensive, make statements about preferred practices, and would meet the definition of a guideline (described in chapter 7) (376). OMAR, however, states that Consensus Development Conference Statements are not intended to serve as guidelines and are often issued during initial technology diffusion, before a guideline would be developed (236). Consensus Development Conference statements are independent reports of the respective panels, and although they are widely perceived as having a federal imprimature, NIH does not consider them to be official policy statements (862).

### *Topic Selection*

Topics for consensus development conferences may be suggested by one or more of the NIH Institutes, Centers, or Divisions, OMAR, or (less frequently) other government health agencies, Congress, or the public (862). OMAR receives about 6 to 12 suggestions per year. The final selection of a topic is made when there is agreement between the sponsoring group within NIH (e.g., one of the Institutes) and OMAR. Among the factors considered by staff when selecting a topic are (237,862):

- public health importance,
- controversy over scientific aspects of the issue,
- availability of evidence cm which to base evaluation of the issues,

[5] AHCPR held a meeting in mid-June 1994 to consider the timing of the update of the cataract guideline (59 FR 24702).

**TABLE C-2: National Institutes of Health Consensus Development Conferences, 1990-94**

| Topic | Date | Sponsoring institute |
|---|---|---|
| Acoustic neuroma | December 1991 | NINDS |
| Adjuvant therapy for patients with colon and rectum cancer | April 1990 | NCI |
| Clinical use of botulinum toxin | November 1990 | NINDS |
| Diagnosis and management of asymptomatic primary hyperparathyroidism | October 1990 | NIDDK |
| Diagnosis and treatment of early melanoma | January 1992 | NCI |
| Early identification of hearing Impairment in Infants and young children | March 1993 | NIDCD |
| Effect of antenatal corticosteriods on perinatal outcomes | February-March 1994 | NICHD |
| Gallstones and laparoscopic cholecystectomy | September 1992 | NIDDK |
| Gastrointestinal surgery for severe obesity | March 1991 | NIDDK |
| *Helicobacter pylori* in peptic ulcer disease | February 1994 | NIDDK |
| Impotence | December 1992 | NIDDK |
| Intravenous immunoglobulln prevention and treatment of disease | May 1990 | NIAID |
| Morbidity and morality of dialysis | November 1993 | NIDDK |
| Noise and hearing loss | January 1990 | NIDCD |
| Primary treatment of ovarian cancer | April 1994 | NCI |
| Recognition and treatment of depression in later life | November 1991 | NIMH |
| Surgery for epilepsy | March 1990 | NINDS |
| Treatment for panic disorder | September 1991 | NIMH |
| Treatment of early-stage breast cancer | June 1990 | NCI |
| Treatment of sleep disorders of older people | March 1990 | NIA |
| Triglycerides, HDL, and coronary heart disease | February 1992 | NHLBI |

KEY NC I = National Cancer Institute, HDL=hlgh-density lipoprotein, NHLBI = National Heart Lung and Blood Institute NIA = National Institute on Aging, NIAID - National Institute of Allergy and Infectious Diseases NICHD - National Institute of Chi l d Health and Human Development, N I DDK = National I nstitute of Diabetes and Digestive and Kidney D seases NIDCD = National Institute on Deafness and other Communication Disorders NIMH = National Insttute of Mental Health NINDS - Nat onal Institute of Neurological Disorders and Stroke

SOURCE Off Ice of Technology Assessment, 1994 based on information from U S Department of Health Human Services Pubic Health Service National Institutes of Health Office of Medical Applications and Research

- amenability to clarification on technical grounds (recommendations should not depend mainly on the impressions or value judgments of panelists), and
- the gap between clinician knowledge and practice.

Other factors considered include public interest and the potential impacts on prevention and cost. The timing of the conference is intended to be nei - ther so early in the developmental course of a new technology that data are insufficient, nor so late

that the conference merely reiterates a consensus already reached by the profession (862).

### Scope of Statements

Consensus conferences may examine either emerging or established technologies. Guideline topics can be either condition specific (e.g., impotence, melanoma, panic disorder) or technology specific (e.g., dialysis, antenatal corticosteroids). The primary focus of consensus conferences is medical safety and effectiveness. but other aspects of a technology may also be considered (e.g., economic, sociologic, legal, and ethical issues) (862).

A planning committee made up of two to three nongovernment researchers, an OMAR staff person, and a staff person from the NIH-sponsoring group(s) identify key questions to be answered at the conference. Usually four to six questions, are posed that cover efficacy, risks, clinical applications, and avenues for future research. For the recent consensus conference on corticosteroids' perinatal effect, for example, the following questions were posed (141 ):

1. For what conditions and purposes are antenatal corticosteroids used, and what is the scientific basis for that use?
2. What are the short- and long-term adverse effects for the infant and mother?
3. What are the economic consequences of this treatment?
4. What is the influence of type of corticosteroid, dosage, timing, circumstances of administration, and associated therapy on treatment outcome?
5. What are the recommendations for use of antenatal corticosteroids?
6. What research is needed to guide clinical care?

### Collecting and Synthesizing Evidence

Panelists receive background materials that include published papers, abstracts of the conference speakers' presentations, and a bibliography prepared by the planning committee. The planning committee sometimes requests that a background review paper or recta-analysis of available literature be conducted. Generally. however, extensive literature searches arc not conducted and evidence is not formally reviewed or synthesized (237). As many as 20 to 30 experts identified by the planning committee are invited to present evidence at the conference.

### Characteristics of Group Membership and Processes[6]

OMAR panels have varied in size from nine to 16 members, averaging about 12 to 13 members. Panelists represent research investigators, health professionals who are users of the technology, methodologists (e.g., biostatisticians and epidemiologists), and representatives of the public and other relevant perspectives (e.g., ethicists, lawyers, patient groups).[7] A nationally recognized expert in the general field under consideration is chosen as panel chairperson. According to OMAR staff, the chairperson is someone who is considered likely to be unbiased and who does not hold any particular advocacy position. The conference planning committee recommends conference panel members and invited speakers.

Consensus conferences usually last three days. During the first day and a half, as many as 20 to 30 experts present information on the state of the science, and data regarding the key questions to be addressed by the panel. The meetings arc open to

[6]This section is based on the following references: 378,545,599,862. Full citations are at the end of this report.

[7]70 MAR decided against balanced panels composed of those representing opposing viewpoints because the a.gene'y felt that strong disagreement among such panelists could have a detrimental effect on the decisionmaking process. Consequently, OMAR seeks a chairperson and panelists who arc neutral (388). To help assure this, the publications of candidate panelists are scrutinized to ensure that they have not published extensively on the conference topic (378). Panelists cannot be federal employees.

the public and include opportunities for questions and answers for all in attendance.

On the evening of the first day, the panel meets in executive session to begin to draft the consensus statement. Subgroups are usually formed to address certain questions. Starting at noon of the second day, the panel again meets in executive session and completes the draft of the statement.

Generally, the group process is informal, but on a few occasions, decision models have been used to help the panel explore the implications of available evidence (386). [8]The panel often works around the clock and under intense pressure to come to agreement. On the morning of the third day, the statement is read publicly at a plenary session and then modified at the discretion of the panel on the basis of comments made by the audience. The statement is then adopted formally by the panel.

Sometimes the process is modified by extending the conferences for an additional day, shortening the time allotted to speakers, holding more than one preliminary panel meeting, or providing the panel with papers or position statements well ahead of the conference (865).

Each panel decides on its own definition of consensus. There is no formal way of assessing level of agreement. Occasionally, if disagreement arises on a particular issue, votes are taken with majority rule. Dissenting opinions are usually resolved by discussion. Only twice have minority opinions been included in the consensus statement (378). The underlying rationale or evidence behind any recommendations or conclusions are usually not included in the consensus statements, nor do most statements include references to the literature that was considered.

The NIH consensus process and its impact on practice have been extensively evaluated and critiqued (372,41 1,5 10,949). Surveys of physicians are sometimes conducted before and after conferences to monitor their impact. However, there have been no formal studies of the reliability and validity of the NIH consensus process. Some new adaptations of the process have been tried and implemented, but the process has remained relatively unchanged in its basic design **(237,600).**

### *Review Process*

The consensus statements do not undergo external review beyond that which occurs at the conference.

### *Updating Statements*

Each year, statements that are 5 years old or older are reviewed. As of late 1993, OMAR had identified 31 out-of-date statements (out of a total of 93 statements). OMAR staff state that they plan followup conferences or amendments for some of these outdated statements (237).

### *Development Time and Cost*

According to OMAR staff, the Consensus Development Conference process general 1 y takes one to one and a half years to complete and costs approximately $150,000.

## National Heart, Lung, and Blood Institute

NHLBI has issued guidelines through three of its educational programs:[9]

- the National High Blood Pressure (HBP) Education Program,
- the National Cholesterol Education Program, and
- the National Asthma and Prevention Education Program.

The purpose of these NHLBI educational programs is to promote the timely transfer of research findings to health professionals, patients, and the

[8]Interestingly, using decision support tools was considered by the panels to be of limited value ( 3781.

[9]A fourth National Education Program on blood resources has been phased out. In addition to its guidelines, NHLBI has made recommendations in several reports regarding the diagnosis and treatment of acute myocardial infarction (454).

general public. The High Blood Pressure Education Program was established in 1972, the Cholesterol Program followed in 1985, and the program on asthma was initiated in 1988.

A defining characteristic of each of the education programs is the establishment of a standing program coordinating committee. The three existing committees are "independent" of the government but are managed by the NHLBI Office of Prevention Education and Control (within the Office of the Director). Coordinating committees meet twice a year and are charged with developing strategies to facilitate the transfer of research findings in their respective areas to clinicians and the public.

Each coordinating committee includes about 35 members who represent professional societies (e.g., American Public Health Association, American College of Cardiology). voluntary health agencies (e.g., American Heart Association), consumer organizations (e.g., Citizens for High Blood Pressure), and government agencies (e.g., AHCPR, Health Care Financing Administration (HCFA)). The coordinating committee may sponsor conferences and workshops, develop patient educational materials, or suggest that guidelines be developed. Subcommittees are formed to address specific issues. A science subcommittee, for example, identifies important emerging scientific issues.

The coordinating committees discuss the merits of guideline development at their meetings and vote to decide whether to proceed with their development. Each of the Education Programs has issued guidelines: five related to HBP, two on cholesterol, and one relating to the management of asthma. The current NHLBI guidelines include:

- Guidelines for the Diagnosis and Management of Asthma (856);
- The Fifth Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood pressure (854); and
- The Second Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (857).

NHLBI is also cosponsoring two guidelines being produced by AHCPR (on unstable angina and cardiac rehabilitation).

### *Topic Selection*

Topics relating to the detection, diagnosis, and management of high blood pressure, cholesterol, and asthma are considered by guideline panels.

### *Scope of the Guidelines*

Each coordinating committee decides on the scope of the guideline. Generally the guidelines are limited to the medical effectiveness and safety of clinical interventions. NHLBI guidelines include patient management protocols in the form of algorithms.

Recently, cost-effectiveness has been considered in some very limited contexts. For example, NHLBI reports that cost-effectiveness studies were used as a basis for targeting interventions to certain groups (e.g., drug treatment versus counseling regarding lifestyle change for hypercholes - terolemia) in the latest guideline of the expert panel on high blood cholesterol (454). This guideline also includes a brief section that discusses cost-effectiveness as a criterion for evaluation and therapy (857). An NHLBI working group is reportedly looking at issues related to the cost-effectiveness of asthma interventions (540).

### *Collecting and Synthesizing Evidence*

Individual members or subcommittee members are generally responsible for the literature review

[10] In 1972, Congress authorized NHLBI to provide the public and the health professions with health information with regard to cardiovascular and pulmonary diseases. Special emphasis was placed on dissemination of information regarding diet, exercise, stress, hypertension, cigarette smoking, weight control, and other factors affecting the prevention of arteriosclerosis and other cardiovascular diseases and pulmonary disease (Public Law 92-423).

for their section of the guideline, so approaches used to identify and synthesize information may vary for different sections of the guideline. Formal methods to categorize, grade, or rank the type of evidence being considered have not been used. Meta-analyses have been used in some recent guidelines (e.g., to review the literature on the effects of cholesterol reduction in people with coronary heart disease) (857).

### *Characteristics of Group Members and Processes*

NHLBI guideline panels are very large, comprising 20 to 50 members. Members of the coordinating committee recommend representatives of their organizations (sometimes themselves) or recognized experts to serve on the panel (most panelists are outside experts). The director of NHLBI, who is also the chairman of the coordinating committee, selects the panel members. Panels include a variety of physicians (both primary care physicians and specialists) and other health professionals (e.g., health educators, nurses, nutritionists). Methodologists such as epidemiologists and economists also sometimes serve on panels. NHLBI staff provide technical and administrative support (540).

Panel subcommittees are formed to address specific aspects of the guideline topic. There are usually five or more meetings of the full panel, as well as separate subcommittee meetings and conference calls. Drafts of guidelines are provided to the coordinating committee for input and are discussed at panel meetings. Final drafts are submitted to the full coordinating committee for approval. Here, voting is used and a majority is required for approval of the guideline. Generally, any problems that the coordinating committee has with the guidelines are ironed out with the panel and guidelines are approved by unanimous vote. If there were serious disagreements among panelists that could not be resolved, a minority report could be issued (although this has not yet occurred) (540).

NHLBI is considering alternative approaches to guideline development. At a recent workshop, the strengths and weaknesses of their relatively informal and flexible approach versus a more standardized structured approach were discussed (855).

### *Review Process*

The guideline is reviewed by members of both the panel and the coordinating committee, who represent a variety of provider and consumer groups. Sometimes guidelines are also reviewed by other outside experts.

### *Development Time and Cost*

According to NHLBI staff. it usually takes about 18 months to complete an NHLBI-sponsored guideline at a cost of about $200,000. (This estimate excludes NHLBI staff-associated costs and publication and dissemination costs. Sometimes the latter costs are assumed by professional societies or other groups (855). )

### *Updating Guidelines*

The standing coordinating committee monitors new developments and determines when a guideline needs to be updated. Over the 22-year period that the HBP committee has been activc, the report on HBP has been updated four times. The cholesterol guideline has been updated once, and there are no immediate plans to update the recently issued asthma guidelines.

## National Cancer Institute

In the last several years, the National Cancer Institute (NCI) has been moving away from issuing guidelines that contain directive recommendations. NCI advisory groups recommended in 1987 that NCI not issue guidelines, but instead issue science-based statements (305). From 1987 to 1993 NCI issued “working” guidelines. In the fall of 1993, in the wake of a debate surrounding breast cancer screening recommendations. NCI began to issue scientific statements instead of guidelines or recommendations (305).

Under the new policy, information to aid in clinical decisionmaking is disseminated to clinicians and patients through NCI’s computerized

PDQ (Physician Data Query) system.[11] Written statements such as brochures and articles are also available through the International Cancer Information Center. NCI statements include brief reviews of available epidemiologic data on a topic without any clinical recommendations. The PDQ includes five categories of cancer information (359):

1. treatment of adult cancer,
2. treatment of childhood cancer,
3. supportive care for cancer patients (e.g., managing pain and nausea),
4. cancer screening and prevention, and
5. investigational and newly approved drugs.

### *Topic Selection*

Information related to screening and prevention, treatment, supportive care, and new anticancer agents are considered for inclusion in the PDQ system by the PDQ editorial boards (see below) (359).

### *Scope of Statements*

PDQ statements include information on efficacy but generally exclude evidence of cost or cost-effectiveness from consideration. Quality of life issues are addressed in the supportive care file (e.g., cancer-related pain, nausea) (359).

### *Collecting and Synthesizing Evidence*

Each month, NCI staff review the tables of contents of more than 70 biomedical journals to identify articles of potential relevance. Articles are retrieved and then screened for relevance and scientific validity. Selected articles are referred to the appropriate editorial board members for review.

Members of the screening and prevention editorial board rate the articles using the following levels of strength of evidence:

1. Evidence obtained from at least one randomized controlled trial;
2. Evidence obtained from controlled trials without randomization;
3. Evidence obtained from cohort or case-control analytic studies, preferably from more than one center or research group;
4. Evidence obtained from multiple time series with or without intervention; and
5. Opinions of respected authorities based on clinical experience, and reports of expert committees.

When rating evidence for primary prevention, the reported outcome-either death, the prevention of metastatic disease, or an accepted validated intermediate endpoint (e.g., large adenomatous polyps for colorectal cancer)— is included in the rating as an "A,'* "B," or "C." For example, a randomizcd clinical trial that considered death as the primary outcome would be rated as 1A. A similar system of evidence rating is currently being developed for use by the adult treatment editorial board (359).

### *Characteristics of Group Members and Processes*

The information in PDQ is updated monthly by five editorial boards. Each board has approximately 10 to 20 members. Some are NCI staff, but most are experts from outside of the federal government.[12] Board members are generally cancer specialists (medical oncology, oncology nursing, radiotherapy, surgery), but methodologists are also included on the screening and prevention board (e.g., statisticians, epidemiologists).

The editorial boards are charged with keeping the PDQ databases up to date. Members are sent relevant literature to review and decide whether it should be incorporated into the PDQ statements, just referenced, or ignored. Judgments about the literature are discussed by each editorial board at

[11] PDQ also includes descriptions of clinical trials that are open or approved for patient accrual, directories of physicians and organizations that specialize in cancer care and screening, and summaries of clinical trial protocols that have been completed or are no longer accepting patients. All protocols that NCI supports are listed in PDQ. Other protocols are included after review and approval by the PDQ editorial boards.

[12] Over three-fourths of the 65 members on the five editorial boards (78 percent) are experts from outside the federal government (359).

monthly meetings when updates to the statements within that board's domain are considered. Group process methods are informal and when there is disagreement, the statement is sent out for review to advisory board members and other experts as appropriate. PDQ statements will reflect uncertainty if there is disagreement regarding the interpretation of evidence. PDQ statements are evidence-based, but the opinion of experts maybe used when more rigorously tested evidence is lacking.

### *Review Process*

An advisory board of over 100 cancer specialists. most from outside the government, also regularly review information statements and suggest updates and changes.

### *Development Time and Cost, and Updating*

PDQ statements are being continually reviewed and revised. so their cost is difficult to assess. The initial content of a statement is developed over a six- to 12-month period. [13]

## CDC Advisory Committee on Immunization Practices

The CDC publishes recommendations and guidelines on numerous public health topics in the *Morbidity and Mortality Weekly Report (75* 1 ). [14] It has also recentl y created a prevention guidelines database that includes about 300 CDC-approved recommendations and guidelines issued since 1982 (224). [15] While many of CDC's recommendations and guidelines relate to gencral public health issues such as surveillance activities and laboratory practices, others are directed to practicing clinicians. Examples of recent CDC clinical practice recommendations and guidelines include:

- General Recommendations on Immunization (837)*
- Recommendations for the Prevention and Man - agement of Chlamydia Trachomatis Infections, 1993 (834),
- Sexually Transmitted Diseases Treatment Guidelines (836).
- Guidelines for Adolescent Preventive Services (830). and
- Standards for Pediatric Immunization Practices (829).

Individual groups within CDC develop recommendations and guidelines using different procedures. Some use standing committees of experts. appointed by the Secretary of Health and Human Services, while others appoint ad hoc panels directly. Sometimes CDC committees formally collaborate with outside groups such as the American Lung Association or the American Diabetes Association (752). The focus of the recommendations and guidelines are often limited to safety and effectiveness, but CDC staff report that panels are increasingly considering including cost analyses in their deliberations. CDC is developing an internal resource guide on decision and economic analysis to facilitate such considerations (831).

An example of a well-established CDC guidelines development activity is the Advisory Committee on Immunization Practices (ACIP), through which CDC sponsors the development of

[13] One board meets monthly (the adult treatment board), while others meet bimonthly (e.g., the prevention board) (359).

[14] CDC also publishes recommendations and guidelines in journals, books, and manuals, and in publications created in conjunction with other organizations (709). CDC does not formally distinguish recommendations and guidelines (752).

[15] The database is currently limited to CDC-generated recommendations and guidelines, but it might in the future include guidelines of other federal or outside groups (752). Not all of the guidelines are clinical practice guidelines. Many are recommendations or guidelines for public health practices such as surveillance and laboratory work. The database is available to federal, state, and local health officials through the CDC's online Wonder program. The database was created jointly by CDC's Epidemiology Program Office, Public Health Practice Program Office, and the Information Resources Management Office (751).

**TABLE C-3: Recommendations of CDC Advisory Committee on Immunization Practices, 1990-January 1994**

| ACIP Guidelines | Release date |
|---|---|
| Diptheria, Tetanus, Pertussis Recommendations for vaccine Use and Other Preventive Measures | August 1991 |
| General Recommendations on Immunization | January 1994 |
| *Haemophilus* B Conjugate Vaccines and a Combined Diptheria, Tetanus, Pertussis, and *Haemophilus* B Vaccine | September 1993 |
| *Haemophilus* B Conjugate Vaccines for Prevention of *Haemophilus* Influenza Type B Disease Among Infants and Children Two Months of Age and Older | January 1991 |
| Hepatitis B Virus A Comprehensive Strategy for Eliminating Transmission in the United States Through Universal Childhood Vaccination | November 1991 |
| Inactivated Japanese Encephalitis Virus Vaccine | January 1993 |
| Pertussis Vaccination Acellular Pertussis Vaccine for Reinforcing and Booster Use—Supplementary ACIP Statement | February 1992 |
| Pertussis Vaccination Acellular Pertussis Vaccine for the Fourth and Fifth Doses of the DTP Series—Update to Supplementary ACIP Statement. Recommendations of the ACIP | October 1992 |
| Prevention and Control of Influenza Part 1, Vaccines | May 1993 |
| Protection Against Viral Hepatitis | February 1990 |
| Rabies Prevention—United States, 1991 | March 1991 |
| Typhoid immunization | July 1990 |
| Update on Adult Immunization | November 1991 |
| Use of Vaccine and Immune Globulins in Persons with Altered Immunocompetence | April 1993 |
| Vaccinia (Smallpox) Vaccome | December 1991 |
| Yellow Fever Vaccine | May 1990 |

SOURCE Off Ice of Technology Assessment, 1994, based on Informallon from U S Department of Health and Human Services, Public Health Service Centers for Disease Control, Advisory Committee on Immunization Practices

immunization recommendations.[16] The Committee issues about eight to 10 guidelines a year (recommendations made since 1990 are shown in table C-3). Most recently the ACIP issued general recommendations on immunizations and on the use of *Haemophilus b* conjugate vaccines and a combined diphtheria, tetanus, pertussis, and *Haemophilus b* vaccine (835). It is currently considering the types and schedules of pediatric vaccines to be purchased and administered under the government's new ● 'Vaccines for Children" Program (58 FR 65725).

### *Topic Selection*

Topics are chosen by the Committee. Developments at the Food and Drug Administration (FDA) (e.g., a manufacturer's submission of a product for regulatory approval) may prompt a review of a new vaccine. A CDC-wide memo is

[16]The Advisory Committee o, Immunization Practices employs one of the more formal processes used within CDC to develop guide lines and recommendations. Other groups within CDC often use less formal methods (709).

circulated periodically to elicit topics for consideration (317).

### *Scope of Recommendations*

Safety and effectiveness are major considerations, but cost effectiveness is also sometimes considered. For example, a formal cost-effectiveness study was funded to support a recent recommendation on chickenpox vaccines (317).

### *Collection and Synthesis of Evidence*

CDC staff develop background materials for the Committee. These materials include published and unpublished literature as well as data provided by vaccine manufacturers. Sometimes invited experts provide additional information at ACIP meetings. Forma] methods to rate and synthesize evidence are not generally used, although recta-analysis has recently been used to assist in updating the recommendations on a tuberculosis vaccine (the BCG, or bacillus Calmette-Guerin, vaccine) (317).

### *Characteristics of Group Membership and Processes*

The standing committee comprises 10 members approved by the Secretary of DHHS and two ex-officio members from the FDA and NIH. [17] Members include pediatricians, infectious disease specialists, and representatives from state health agencies. Liaison representatives from various provider groups (e.g., American Academy of Pediatrics. American Academy of Family Physicians ), the Departments of Defense and Veterans Affairs, and other professional and advisory groups also attend committee meetings (e. g., Hospital infections Control Practices Advisory Committee, National Vaccine Program).

Committee members consider and discuss evidence at their meetings, which are held three times a year and last one and a half days each. Usually, CDC staff draft the rccommendations, which are reviewed and revised at subsequent meetings until they arc approved. On occasion, the ACIP chairperson may appoint committee members to serve on a working group to formulate a draft recommendation. A recommendation is considered approved when a majority of members approve it: unanimity is not required (317).

### *Review Process*

External review of recommendations is not routine. It occurs occasionally when additional specialized expertise is needed that resides outside of CDC and the committee (709).

### *Development Time and Costs*

Information on development time and costs for ACIP guidelines was not available.

### *Updating Guidelines*

The Committee periodically reviews guidelines and updates them. Influenza vaccine recommendations, for example, are routinely reviewed annually because influenza viral strains change from year to year. Other recommendations are updated when new data or new technologies indicate the need for a change (e.g., the development of a chickenpox vaccine) (709).

## Office of Disease Prevention and Health Promotion: US. Preventive Services Task Force

The U.S. Preventive Services Task Force (USPSTF) was created in 1984 by the Office of Disease Prevention and Health Promotion (within the Office of the Assistant Secretary for Health) to develop evidence-based practice guidelines for preventive care, following the model set by the Canadian Task Force on the Periodic Health Examination (100,507). USPSTF members have reviewed evidence of the effectiveness of 169 preventive services for the prevention of 60 target conditions and have made age-. sex-, and risk factor-specific recommendations in the *Guide to*

[17] Members are generally appointed to serve four-year terms (432).

*Clinical Preventive Services,* published in 1989 (871).

A new panel began work in 1990 to update previous recommendations and to evaluate preventive services not examined by the first panel (7 17). An updated Guide will be published in late 1994, covering an even broader range of preventive services (947). Some of these recommendations have been published recently in the *Journal of the American Medical Association* (JAMA) (e.g., those relating to screening for adolescent idiopathic scoliosis (872), home uterine activity monitoring for preterm labor (873). and routine iron supplementation during pregnancy (874)). As with most other government-sponsored guideline panels, the USPSTF is independent and recommendations are not required to pass through official government clearance.

### *Topic Selection and Scope of Guidelines*

The USPSTF develops evidence-based guidelines on preventive services offered to asymptomatic individuals by primary care providers. ] [9]The Task Force provides recommendations on the appropriate delivery of these services in the periodic health examination. Preventive services include screening tests, immunizations, chemoprophylaxis, and patient counseling. The Task Force considers the efficacy, effectiveness, safety, appropriateness, and costs of services. It does not conduct formal cost-effectiveness studies, nor does it focus on issues such as barriers to implementing a recommended service (e.g., lack of reimbursement for preventive services). Target conditions are identified and then the range of clinical preventive services that might be effective in preventing the condition are described (378). Topics are selected for review based on (948):

- the severity and frequency of the target condition (burden of suffering),
- uncertainty about appropriate practice that can be remedied by guidelines,
- timeliness of the topic,
- costs,
- availability of scientific evidence, and
- feasibility of the review.

### *Collection and Synthesis of Evidence*

The Task Force relics on formal criteria of effectiveness and grades the quality of individual studies according to epidemiological principles. A hierarchical system is used to rate studies on the basis of the study design and methods (see box 7-2, chapter 7). Usually, only published peer-reviewed data are considered. Expert opinion may be considered, but it is given a different rating than empirical data. Evidence is summarized and published in tabular format. Recommendations of the panels are linked directly to evidence. Where data are lacking, recommendations are often presented in flexible or neutral language (e.g., “insufficient evidence to recommend for or against”). The panels have frequently used formal methods of information synthesis such as meta-analysis and decision analysis (948).

The Task Force has also adopted explicit methods for organizing the review of evidence, such as using “causal pathways” to frame the evaluation of evidence (44,717). If evidence is lacking on the association between a preventive service and the outcome of interest, the panel examines evidence along the causal pathway. If there is no evidence on the effect of screening adolescents for idiopathic scoliosis on reducing Scoliosis-related morbidity (e.g.. disability). for example, the panel examines intermediate relationships. These include the relationship between screening and diagnosing scoliosis early, and the relationship between early intervention and subsequent health outcomes, such as back complaints, disability,

[18]The USPSTF also receives outside support from various associations and foundations, such as the Kellogg Foundation and the American College of Preventive Medicine.

[19]The USPSTF does not address preventive services delivered at the worksite, school, or other community settings (717).

and psychosocial effects) (see figure 7-1, chapter 7).

### *Characteristics of Group Membership and Process*

The original USPSTF panel was composed of *20* individuals: 14 primary care physicians, 3 other health care providers, an economist, a medical sociologist. and a health services researchcr. The new panel consists of 10 core members (8 primary care physicians and 2 methodologists) and liaisons from primary care specialty societies, U.S. government agencies, and the Canadian Task Force on the Periodic Health Examination (948).[20] A systematic method was used to select task force members. A panel of senior advisors, including former members of the USPSTF regularly provides consultation. Staff of thc DHHS Office of Disease Prevention and Health Promotion provide administrative support for the activities of the task force.

No formal group consensus development methods are currently used beyond simple voting, and when data are unavailable the task force does not attempt to use the opinions of group members as a basis for making recommendations. The recommendations are strictly science-based and the rationale for each one is documented in an explicit format. The documentation includes a description of the evidence, with complete citations, and a detailed explanation of how the evidence was interpreted. The criteria of clinical effectiveness vary depending on the type of preventive service. An evaluation of a screening test, for example, considers the test's accuracy and reliability, and the effectiveness of early detection in improving health outcomes. An evaluation of a counseling intervention would consider information on the effectiveness of behavior change on risk reduction and health outcomes (948). A grade is assigned to each recommendation representing the strength of the supporting evidence (see box 7-2, chapter 7).

Differences in interpretation of the evidence are discussed at task force meetings. To date, there have been no dissenting opinions, but if there were. they would be documented in the relevant report. Prevention recommendations of other groups are published alongside the panel's recommendations. The USPSTF has sometimes endorsed another group's recommendations after some independent examination of the evidence (947) (e.g., the American College of Physicians' (ACP) recommendations on hormone replacement therapy and on screening for ovarian cancer) (926).

### *Review Process*

Draft guidelines are extensively reviewed by experts in the relevant topics in the United States. Canada, and Europe (871).

### *Development Time and Cost*

Estimates are not available.

### *Updating* Guidelines

The mission of the new USPSTF panel is to update previous recommendations and issue a revised edition of the *Guide to Clinical Preventive Services (717). New* scientific evidence is examined systematically on a periodic basis to identify recommendations that require reevaluation.

## PRIVATE GUIDELINE ACTIVITIES

Private guideline efforts abound and include those of physician organizations (e.g., American Medical Association), voluntary organizations (e.g., American Cancer Society). health care organizations (e.g., Harvard Community Health Plan), and research organizations (e. g., RAND).

Guidelines produced by physician organizations are especially prominent among private guideline development activities. Among the first guidelines written by a physician organization was the American Academy of Pediatrics' 1938

[20]Representatives of these groups regularly attend the USPSTF meetings and review draft recommendations (717).

monograph on infectious disease control (376). By 1993, about 50 physician organizations were involved in related efforts, contributing to the development of about 250 to 300 new guidelines each year **(8 15).** cSometimes specialty groups **work independently** to develop guidelines for **their members (e.g.,** the American Academy of **Ophthalmology), whi**le other groups create guidelines that are intended to be used more broadly across specialties (e.g., the ACP). Reasons for developing guidelines and methods used to develop them vary widely among the physician groups that are creating them (880).[21]

This section summarizes the processes used by a few selected groups that have well-established guideline activities, including the activities of two physician groups (the American College of Physicians (ACP)) and the American Medical Association), a health maintenance organization (Harvard Community Health Plan), and a research group (RAND Corporation). Examining the different approaches taken provides a broader context with which to review federal guideline efforts.

## The American College of Physicians

ACP is the largest physician specialty society, with a membership of about 80,000 internists (925). ACP began developing guidelines in 1981 through its Clinical Efficacy Assessment Project (CEAP) and since then has developed more than 160 assessments to guide its members' practices (22).[22] The purpose of CEAP guidelines is to provide continuing education, and to improve the efficiency of medical practice by reducing use of unnecessary tests and procedures (378). The ACP views their guidelines as potentially useful in establishing reimbursement policies, utilization and systems management, informing hospital purchasing, and formulating research agendas (378). ACP issues about three to four guidelines per year (925). Some recent examples include:[23]

- Treatment of Gallstones (17),
- * Ambulatory Blood Pressure Monitoring (15),
- Screening Guidelines for Diabetic Retinopathy (18),
- Practice Strategies for Elective Red Blood Cell Transfusion (14), and
- Preventive Care Guidelines: 1991 (332).

ACP has developed a mechanism to approve the guidelines of other groups. For example, the USPSTF guideline on screening for genital herpes was recently formally endorsed by ACP (925).

### Topic Selection

A Clinical Efficacy Assessment subcommittee[24] identifies technologies that are potential candidates for assessment based on surveys of ACP members. Final decisions on topics are made by the subcommittee using six criteria:[25]

---

[21] In 1990, the U.S. General Accounting Office (GAO) interviewed representatives from 27 of the 35 medical specialty societies that had been identified as possessing or developing guidelines. During interview's GAO determined: why guidelines were developed; what kinds of guidelines were developed (scope, types of recommendations, types of products); the methodology used to develop guidelines; who was involved in dev eloping guidelines; how guidelines were disseminated: what provisions existed for updating guidelines; and how much effort was required to produce guidelines (880). The details of GAO's findings are not presented in its report due to confidentiality issues.

[22] The ACP first became involved in guideline development in the mid-1970s when it assisted the national Blue Cross and Blue Shield Association in determining whether selected medical procedures were outdated, had been replaced, or were not effective (380)(see chapter 6).

[23] Recent ACP reviews and guidelines are compiled in *Clinical Practice Guidelines (16).* Other guidelines appear in three ACP publications, *Common Diagnostic Tests, Common Screening Tests,* and *The Guide for Adult Immunization (925).*

[24] The Clinical Efficacy Assessment Subcommittee is a standing committee of the Health and Public Policy Committee of the ACP.

[25] Technologies that are potential candidates for assessment are identified through a number Of sources. Internal sources include subcommittee reviews of policy needs, practitioner opinion, academic opinion, recent journal articles, ACP committees, requests by outside organizations (e.g.. government agencies and third-party payers), and professional meetings. External methods of identification include recommendations and requests from ACP members and surveying members regarding topics for guidelines. The surveys began in 1993, and ACP plans to continue them on a routine basis (925).

- burden of illness,
- clinical impact of technology,
- aggregate costs associated with the technology,
- relevance of the technology to internists,
- practicing physicians' degree of uncertainty regarding appropriate use of the technology,
- adequacy of the knowledge base for an assessment, and
- likelihood that an assessment will result in altered practice patterns.

A weighting and ranking method (adapted from one that has been used at AHCPR) has been adopted to guide priority setting (925). ACP publishes a notice of assessment in the *Annals of Internal Medicine* and the *ACP Observer.* Readers are invited to submit written comments and data.

### Scope of Guidelines

CEAP evaluates drugs, specific applications of medical technologies, surgical procedures, laboratory tests, and management strategies (e.g., how to study the gallbladder). Assessments are comprehensive and can include considerations of screening, prevention, diagnosis, treatment, and rehabilitation. Topics of new assessments will probably be more condition-oriented (925). Assessments routinely compare alternative techniques. Assessments may focus on new, established, or obsolete technologies and practices (378). CEAP considers evidence of cost-effectiveness, but formal cost-effectiveness studies are restricted to guidelines on screening for disease in asymptomatic persons (925).

### Collecting and Synthesizing Evidence

ACP uses an evidence-based approach, relying to the fullest extent possible upon literature rather than expert opinion. Expert consultants summarize relevant literature in a background paper for the guideline panel.[26] CEAP subcommittee members who are not generally experts in the area under consideration also review relevant literature. Levels of evidence similar to those used by the USPSTF are used to rate the literature reviewed for the guideline (378).

The rating and interpretation of each paper is discussed at group meetings. The final background paper, including recommendations, is structured to include (13):

- a background review;
- an explicit statement of objectives;
- a description of the methods of analyzing published data, including data-search specification and criteria used for accepting or excluding studies;
- a description of quantitative methods (e.g., Bayesian analysis, decision trees, cost-effectiveness analyses) and tables showing key variables and data from various studies;
- a comparison of alternative technologies, where applicable;
- a presentation of findings, including the level of evidence for or against positions and the explication of circumstances of marginal benefit compared with similar techniques; and
- a summary table of recommendations.

Authors are encouraged to identify and use state-of-the-art methods for performing secondary data analysis, and statements of recommendations must conform to detailed requirements (e.g., levels of evidence must be documented). The final background paper includes the extensive literature review with assigned ratings of levels of evidence. The paragraphs of the paper are numbered and are used as reference numbers for the shorter clinical guideline. This allows readers to easily go back to the evidence supporting any recommendation of the guideline.

### Characteristics of Group Members and Processes

Guidelines are developed by a standing committee of eight physicians representing research methodologists, practicing general internists, and

[26]One key author is identified to complete the background paper. Often the key author relies on one to five other experts to develop the paper (925).

subspecialists.[27] This guideline panel is assisted by one or more expert consultants on each guideline topic. The expert consultants are responsible for drafting an extensive review of the literature and recommendations for subcommittee consideration. The consultant(s) and subcommittee members meet five times a year (each meeting lasts one to one and a half days). Meetings are regularly attended by representatives of USPSTF, AHCPR, and medical specialty societies.

The background paper completed by the consultants lays the foundation for the guideline. The recommendations and the strength of the evidence underlying the recommendations are discussed at length at scheduled meetings. The subcommittee makes suggested changes to the background paper and consultants generally rewrite a draft several times. Once a guideline draft is approved by the subcommittee, it is sent out for review. Comments are then considered by the subcommittee. Structured group process methods are not used. Votes are sometimes taken to settle disagreements, but consensus is almost always reached by the subcommittee and its consultants. Once the recommendations are formulated, each is assigned a grade from "A" to "C" according to the level of evidence available to support it (378):

The ACP plans to strengthen its CEAP program by (378):

- using new methods for assessing data, including consideration of patient preferences,
- revising formats for guidelines,
- making draft guidelines avail able online for a network of members who will pretest the guidelines and then measure patient outcomes when the guidelines are used according to specific protocols,
- starting a formal convening activity to involve multidisciplinary groups in the development of guidelines, and
- developing a systematic and perhaps new way of updating guidelines.

Differences in opinion between expert consultants and the subcommittee are rare (they have occurred twice), but when they occur, they are acknowledged in the final paper. For example, some differences in interpretation of the evidence regarding the use of automated and patient blood pressure devices were acknowledged in a recent ACP guideline (15).

### Review Process

Draft guidelines are extensively reviewed. Medical societies, manufacturers, researchers, and others identified as "stakeholders" in the guideline recommendations are asked to review the drafts. Reviewer comments are considered by the consultants and the CEAP subcommittee, and drafts are revised as appropriate to CEAP's mission to be "evidence-based" (925). Once the guideline is approved by ACP, it is submitted to the *Annals of Internal Medicine* for publication.

### *Development Time and Cost*

Guidelines are generally completed within one to two years at a cost of $30,000 to $50,000 each (926).

### *Updating Guidelines*

ACP guidelines are reviewed annually, with new guidelines issued as necessary when new evidence becomes available. Current CEAP guidelines have been published in a compendium of ACP-approved practice guidelines (16).

---

[27] Subcommittee members serve one-year terms that are renewable up to five times (925). Members must adhere to conflict of interest policies that were adopted by ACP in July 1993 (925).

## American Medical Association

The American Medical Association (AMA) represents approximately 297.000 physicians and 82 medical specialty societies (407).[28] The AMA has assumed a coordination role in guideline development[29] through its Forum on Practice Parameters and its Practice Parameter Partnership[30]. The AMA has also developed a system to track the development, publication, and withdrawal of guidelines. Information from the tracking system is published in the *Practice Parameter Update.*[31]

Since **1982,** AMA has also developed its own practice recommendations through its Diagnostic and Therapeutic Technology Assessment program (DATTA). In this program, a select group of practicing physicians are sent a literature review and polled regarding the safety and effectiveness of a particular technology. Unlike most other guideline efforts, results arc based on survey results rather than the combined judgment of a group that meets face-to-face. The DATTA survey results are published in *JAMA.* DATTA assessments published since 1993 include:

- Lung Transplantation (530).
- Teflon Preparations for Urinary Incontinence (531 ),
- Human Papillomavirus DNA Testing in the Management of Cervical Neoplasia ( 148), and
- Hyperthermia As Adjuvant Treatment for Recurrent Breast Cancer and Primary Malignant Glioma (532).

The DATTA program evaluates the safety and effectiveness of drugs, devices, and procedures. New, established. and potentially obsolete technologies are reviewed. Costs are not considered. As of 1993, 72 DATTA evaluations have been completed (19).

### Topic Selection

Criteria for selecting topics include (20):

- existence of controversy or large uncertainty about the technology in the medical community.
- potential for the technology to affect large numbers of patients and/or contribute to substantial costs/cost-savings,
- existence of available data on the technology,
- potential for the evaluation to benefit physician practice and improve patient outcome, and
- potential for the assessment to have an impact-e. g., to affect the diffusion of a promising technology or protect patients from a possible fraudulent technology (20).

Questions for DATTA evaluations are considered from a variety of sources (e.g., physicians, patients, third-party payers, peer reviewers). Each year a survey to identify technologies for DATTA evaluation is sent to DATTA subscribers (these include medical directors of HMOs and other health care facilities, third-party payers, and benefits consultants). Physicians on the DATTA reference panel and medical specialty societies also may be

[28] Nonmember physicians who belong to specialty societies in the AMA House of Delegates are also represented by the AMA (407).

[29] The AMA defines practice parameters as "strategies for patient management, developed to assist physicians in clinical decision making. These practice parameters include standards, guidelines, and other patient management strategies" (22).

[30] The Forum on Practice Parameters involves more than 80 physician organizations, including specialty and state medical societies (21). Its purpose is to help physician organizations share information on guideline activities and to explore strategies to improve the quality of practice parameters ( 22 ). The Practice Parameters Partnership is a smaller group of organizations that make policy decisions regarding practice parameters. It includes AHCPR and 16 large specialty societies (2! ). Its mission is to direct and influence the &J elopment, implementation, and application of practice parameters (425). The Partnership has recently reviewed practice parameters to assess their conformance w ith the AMA attributes for guideline development ( I0).

[31] The 1993 edition lists about 15(0) practice parameters developed by more than 45 physician organization and other groups (21).

asked for suggestions. Up to eight evaluations are conducted per year (407).

### Scope of Guidelines

DATTA evaluations are generally limited to considerations of the safety and effectiveness of a given technology.

### Collecting and Synthesizing Evidence

For each DATTA evaluation, an AMA staff member, consultant physician, or consultant medical scientist prepares a literature review. The review includes a description of the methods used to identify the relevant literature. Additional information is sought as needed from manufacturers or prominent researchers in the field. The literature review is reviewed externally by physicians nominated by relevant specialty societies. Comments of the reviewers are incorporated into the final paper.

### Characteristics of Group Members and Processes

A panel of at least 20 physicians is selected from a database of experts maintained by the AMA. The database includes a listing of over 2,500 physicians who are nominated by AMA councils, deans of medical schools, state medical societies, and national specialty societies. Panelists need not be members of the AMA, but they must have experience with the technology being evaluated. They may be:

- referring physicians-those who provide care on a regular basis for patients with conditions for which the technology being evaluated is an optional intervention;
- performing physicians—those who perform the technology being evaluated, or a competing technology, currently or in the recent past;
- followup physicians—those who followup patients after the procedure to observe, as relevant, short-term and long-term outcomes of the procedure; and
- researchers —those who conduct clinical, basic or epidemiologic research involving the technology.

Usually, no more than half of panelists perform the technology being evaluated; the remaining half are distributed among referring, followup, and research physicians. Panelists must sign statements indicating that they are free of direct financial conflict of interest (20).

Selected panelists are sent a literature review on the topic for consideration[32] and are asked to rate the safety and effectiveness of the technology using a standard set of definitions as follows:

Established—accepted as appropriate by the practicing medical community for the given indication in the specified patient population;

Promising—given current knowledge, this technology is appropriate for the given indication in the specified patient population;

Investigational---evidence insufficient to determine appropriateness, warrants further study. Use of this technology for given indication in the specified patient population should be confined largely to research protocols;

Doubtful—given current knowledge, this technology is inappropriate for the given indication in the specified patient population; and

Unacceptable—regarded by the practicing medical community as inappropriate for the given indication in the specified patient population .

Panelists may indicate “no opinion” if they do not consider themselves qualified to respond.

Nonrespondents are followed up until response rates are at least 80 percent. A random sample of nonrespondents may be contacted to determine reasons for their nonresponse. Sensitivity analyses are performed when appropriate and reported in each DATTA evaluation.

[32]The original DATTA methodology did not include a literature review provided to panelists. The process was revised to include it in mid- 199 I.

Medians and 95 percent confidence intervals are calculated for both safety and effectiveness ratings. Agreement or consensus among panelists exists if the shape of the distribution of responses among the five response categories is unimodal and it is determined that the responses differ from what would be expected by chance (i.e., if ratings were assigned by chance, each of the five ratings would have an equal chance of being selected—20 percent).

### *Review Process*

DATTA evaluations undergo peer review and are published in *JAMA.*

### *Development Time and Cost*

The DATTA process takes from six months to a year, excluding prepublication review time at *JAMA (407).*

### Updating Evaluations

As new evidence becomes available. DATTA evaluations are reevaluated. To date, eight DATTA topics have been reassessed and updated.

## Harvard Community Health Plan

The Harvard Community Health Plan (HCHP) is a combination staff- and group-model health maintenance organization based in Boston, with over 50 delivery sites and more than 550.000 members (292). Guidelineș ụșually in the form of algorithms, have been developed as part of a quality improvement program since 1986. Clinical algorithms are "logic trees" that set forth step-by-step procedures for making sequential clinical decisions. The general purpose of the HCHP guidelines effort is to decrease practice variation among HCHP clinicians and improve the overall quality of care rendered to patients (22). As of early 1994, algorithms on over 30 clinical topics had been completed or were under development (table C-4) (291). The overall HCHP guideline development process is summarized in box C-1.

### *Topic Selection*

Each year, HCHP clinicians and managers nominate clinical quality improvement projects. The nominations are reviewed by a committee of medical directors who then designate project leaders, project teams, timelines, expected outputs, expected resource allocations. and predetermined measures of success (291 ). Criteria for choosing topics include:

- common clinical condition;
- unexplained variation in clinical practice (perceived or documented);
- unexplained variation in utilization of limited or costly resources;
- unexplained variation in internal or external referral patterns;
- general clinical uncertainty or controversy:
- uncertain indications for risky or costly intervention;
- internal resource access or supply constraints;
- apparent risk management problem:
- introduction of new diagnostic test, therapeutic procedure, or medication; and
- quality of care problem perceived by patients, clinicians, or managers.

Approximately five topics are selected per year.

### Scope of *Guidelines*

HCHP guidelines usually address issues of safety, efficacy, effectiveness, appropriateness, cost, cost-effectiveness, system impact, risk management implications, and implementation (378). Cost is explicitly considered during the guideline development process. Patient preferences as reflected in patient surveys, focus groups, or interviews are incorporated into the guidelines. In addition. some guidelines provide guidance about eliciting patient preferences and basin: decisions on the results of that process (292).

### Collection and Synthesis of the Evidence

Before the first team meeting, the project leader reviews and evaluates the literature, and distributes relevant articles and a first draft or "seed" al-

**TABLE C-4: Algorithms and Guidelines Completed or Under Development by Harvard Community Health Plan, 1993**

| Internal medicine and surgery | Pediatrics and child mental health | Obstetrics and gynecology | Adult mental health |
|---|---|---|---|
| Asthma, acute and chronic | Anorexia nervosa | Antepartum assessment | Alcohol disease |
| Breast lumps | Asthma, chrome | Gestational ,diabetes | Anxiety |
| Carpel tunnel syndrome | Child abuse | Ectopic pregnancy | Clozapine |
| Colon cancer, screening and followup | Headache, acute | Infertility | Day hospital Indications |
| Diabetes, routine care | Otitis media | Intrauterine growth retardation | Depression |
| Dyspepsia | Screening, routine | Pap smear. followup | Elder abuse |
| Dysuria, acute | Sexual abuse | Pelvlc pain | Fluoxetine |
| Gallstone lithotripsy | Substance abuse | Substance abuse in pregnancy | Panic states |
| Headache | Urinary tract infection | | Sexual assault |
| Hypercholesterolemia | Wheezing, acute | | Substance abuse |
| Hypertension | | | Suicide |
| Immunization, routine | | | |
| Lower back pain | | | |
| Lumbar radiculopathy | | | |
| Pyelonephritis | | | |
| Screening, routine | | | |
| Temporal arteritis | | | |
| Thyroid nodules | | | |

SOURCE L K Gottlieb, H N Sokol, K O Murrey, et al "Algorithm-Based Clinical Quality Improvement, " *HMO Practice 6(1 ) 5-12,* 1993

gorithm to the participants for their review. Seed algorithms may be older versions of an algorithm on the same topic, derived from existing texts or articles, or may be constructed "from scratch."[33] The seed algorithm serves as a starting point for group discussion (291 ).

A wide variety of evidence may be considered depending on the nature and complexity of the issue being addressed (e.g., meta-analyses, cost effectiveness analyses, decision analyses). MEDLINE® searches are used to identify published materials. New unpublished data and expert opinion are sometimes considered. Evidence is usually weighted in an informal way with occasional formal classification of the quality of the evidence (378). HCHP also contracts with qualified individuals and groups to undertake more formal analyses as needed. In addition to published evidence, clinical practice data, risk management data, and cost data from HCHP are frequently used.

### *Characteristics of Group Members and Processes*

The guideline project team is multidisciplinary and includes several intended users of the algorithm, representatives from specialties with expertise in the particular area under consideration, and one or more representatives from the departments of pathology, radiology, pharmacy, or the laboratory if relevant to the topic. Issues consid-

[33] Increasingly HCHP has adapted guidelines developed by national groups for local implementation rather than developing them 'rem scratch. For instance, HCHP has adapted NHLB1'S recently issued asthma guidelines (292).

**BOX C-1: Harvard Community Health Plan Clinical Improvement Process**

I Project definition and organization
- A List and prioritize problems
- B Define project and team

II Conceptual design Clinical guideline development
- A Identify relevant Individuals and assess their needs
- B Develop consensus guideline

III Problem prevention and Implementation
- A Consider potential problems and causes
- B Develop support systems for prevention
- C Design measurement systems
- D Implement new processes

IV Holding the gains Measurement and evaluation
- A Measure performance process and outcomes
- B Monitor systems

SOURCE L K Gottlieb H N Sokol K O Murrey et al "Algorithm-Based Clinical Quality Improvement " *HMO Practice* 6(1) 5-12 1993

ered in making up the team include the clinical specialty or specialties for which the algorithm is intended, and the training level of clinicians or administrators that will be expected to use the algorithm (e. g., physicians, nurses, advanced-practice nurses). HCHP considers the ideal team size to be eight to 12, with all teams having a minimum of five members and a maximum of 15. The project team leader is usually a program coordinator from the HCHP Clinical Guidelines Program staff from the medical director's office (291).[34]

The guideline team first identifies the target patient population. enumerates the desirable clinical outcomes, and assesses the needs of the diverse caregivers. The consensus development process then begins with a brief introductory lecture on al - gorithm construction, nominal group process. and the Delphi method (see chapter 7). Guideline development usually requires three or four 2-hour meetings of the group for discussion, algorithm training, and the actual performance of the nominal group process. This is usually followed by one or two rounds of a Delphi process to reach final consensus. Once the participants have reached consensus on the algorithm, annotations are added in order to clarify or expand on the content of the algorithm, point out remaining areas of controversy, and provide citations to the literature that sup-

[34]In the future, team leaders will be drawn from a wider source of clinical leaders with Clinical Guidelines Program coordinators serving as facilitators and consultants (291).

port the recommendations of the algorithm. The group leader acts as facilitator and in general does not share his or her opinions or vote during the process.

Consensus on a guideline means that every member of the consensus panel can support the guideline (378). If unanimous approval is not achieved, dissenting opinions are included in the final guideline (this has occurred only once) (292).

### *Review Process*

Guidelines are extensively reviewed by HCHP clinicians and managers and are occasional y sent to outside experts for review and comment (292).

### *Development Time and Cost*

The guideline development process generally takes six months and costs approximately $10,000 (excluding the implementation phase) (292).

### *Updating Guidelines*

An "algorithm keeper" is assigned to the guideline and is charged with judging when, or if, clinical advances have rendered the guideline obsolete or in need of revision (628). Criteria for updating guidelines state that the longest review interval is three years, with early review occurring if significant shortcomings of the guideline are discovered after an initial period of use, if important advances in the relevant clinical area occur, or if significant changes in the HCHP delivery system require modification of the guideline (22).

## RAND Corporation

RAND developed a method to rate the appropriateness of indications for medical and surgical procedures as part of the 1984 RAND/UCLA Health Services Utilization Study. The RAND ratings have been used to retrospectively assess the appropriateness of care (as indicated in patient charts). They have also been applied prospectively within precertification programs. Because so many indications are rated for any one procedure (sometimes thousands of indications are rated), the ratings themselves cannot easily be used by practitioners. The method could, however, be adapted to develop practice guidelines (823). RAND has rated appropriateness for the following procedures:

- coronary angiography,
- coronary artery bypass graft surgery,
- carotid endarterectomy,
- percutaneous transluminal coronary angioplasty,
- abdominal aortic aneurysm surgery,
- diagnostic upper gastrointestinal endoscopy,
- colonoscopy,
- cholecystectomy,
- hysterectomy, and
- spinal manipulation for lower back pain.

### *Topic Selection*

RAND selected procedures for evaluation in the Health Services Utilization Study based on the perceived potential of appropriateness criteria to improve the quality of medical care and reduce costs. They considered the number of procedures performed annually, the costs, the risks, and the amount of controversy that exists concerning the appropriateness of use. Recently, RAND assessed the appropriateness of four common and controversial procedures in cooperation with the Academic Medical Center Consortium (AMCC)[35] and the AMA.[36]

[35]The AAMC represents 12 academic medical centers. The four procedures studied were carotid endarterectomy, coronary artery bypass graft surgery, abdominal aortic aneurysm surgery, and cataract surgery (622).

[36]RAND has also developed a clinical Practice guideline (on congestive heart failure) under contract with AHCPR (using the AHCPR guideline development method).

### *Scope of Appropriateness Criteria*

Appropriateness is defined by RAND to mean that a procedure is worth doing if the expected medical benefit to the patient ( health status, quality of life, longevity) exceeds the expected negative consequences to the patient (pain, disability, risk of death). Cost is not explicitly included in the definition of appropriateness (823).

## Collecting and Synthesizing Evidence

Comprehensive background papers are prepared for the RAND panel by physicians with expertise in health services research and epidemiology'. The review begins with a MEDLINE search for all relevant articles about the efficacy, utilization, complications, cost. and stated indications for the procedure of interest. Experts in the field are asked about possible omissions in the reference list. When literature databases are searched. search strategies are documented. Identified literature is classified as original research studies, editorials. reviews, or textbooks. Original research studies that contain primary data are further classified as being (823):

- randomized controlled trials (RCTs),
- prospective non-RCT cohort studies.
- prospective non-RCT registry studies.
- retrospective adjusted cohort and case-control studies,
- observational and unadjusted retrospective cohort studies.
- cross-sectional studies. and
- surveys.

Studies are generally not included if they are case reports.

When possible. scoring systems are used to rate articles. which take into consideration factors that influence reliability and internal and external validity (83,687). Evidence tables are used to present data from the literature (823). Here. complications and effectiveness are shown by clinically homogeneous groups (in so far as possible ). Formal meta-analysis has not been done because, according to RAND, in most cases the data preclude such quantitative analyses (823).

The literature review is used to help panel members develop a list of the clinical circumstances (indications) under which a particular procedure has been shown to be, or is thought to be, beneficial. The number of indications per procedure has varied from as few as 49 for cholecystec - tomy to as many as 3,000 for colonoscopy (823).

### *Characteristics of Group Members and Processes*

A multi specialty group of nationally known clinicians is convened to rate the appropriateness of identified indications. RAND panels have historically consisted of nine members, but the method can be adjusted to include up to 12 (823). The RAND research staff determine the distribution of specialties for the panel. The nine-member coronary artery bypass panel, for example, included one family physician, two internists, three cardiologists, two cardiac surgeons, and one radiologist. Medical societies representing the relevant specialists are asked to nominate five individuals for each of the panel slots. Panel members are selected with an effort made to balance the panel by specialty, geography, and practice type (academic or private practice). Panelists include both those who refer for and those who perform the procedure, but the number of panel members who perform the procedure is four or fewer. Panels are led by physician-researchers, usually the person who has had major responsibility for the literature analysis. The leader is never a person who performs the procedure being evaluated.

Participants are given a literature analysis and a list of indications. Each participating physician is asked to rate each indication on a nine-point appropriateness scale (using the RAND definition of appropriateness) (823). A rating of 1 is a judgment that performance of the procedure for the indication is extremely inappropriate; a rating of 9 is a judgment that the procedure is extremely appropriate. Cost is not considered explicitly. although results of economic studies are included in the literature review.

A modified Delphi group process is used. Panelists perform the first round of' ratings indepen-

dently at home. The ratings from all of the individual physicians are then collated and presented at a meeting of the group, where they are discussed. The ratings are presented anonymously, except that each individual is reminded privately of his or her rating for each indication. Following a structured discussion period, each panelist r-e-rates each indication. No effort is made to reach a consensus on appropriateness ratings. Panelists meet for a total of about two days (823).

Each indication is classified as being appropriate, inappropriate, or equivocal, according to its median rating and the presence of agreement or disagreement among panel members. A rating of 7 to 9 is considered appropriate, 4 to 6 is equivocal. and 1 to 3 is inappropriate. When there is disagreement, the indication is rated as equivocal, irrespective of the median score. Disagreement among raters is defined as at least three ratings in the 1 to 3 range and three ratings in the 7 to 9 rangc.[37] Scores of appropriate and inappropriate are considered to be "With agreement" when, after discarding the one highest rating and the one lowest rating. the remaining ratings are within a three-point range. There is no effort to seek consensus; final ratings are characterized as "with agreemerit. " Typically, fewer than half of the appropriate or inappropriate indications are found to be "with agreement. "

A procedure is considered "necessary" if all four of the following criteria are met (823):

1. the procedure is appropriate,
2. it would be improper not to provide the service,
3. a reasonable chance exists that the procedure will benefit the patients, and
4. the benefit to the patient is not small.

One researcher has suggested an interesting adaptation to the RAND methodology. In addition to rating appropriateness, clinical scenarios could be rated according to whether the evidence is sufficiently inconclusive that it would be ethical to randomize patients in a clinical trial comparing routine use of a technology versus no use (555). Agreement on the acceptability of randomization would provide evidence to support inclusion of patients in multicenter RCTs (555).

### *Review Process*

There is no external process for a review of ratings.

### *Development Time and Cost*

The entire process takes from six months to a year to complete. A 1993 estimate of the cost of each RAND evaluation was $350,000.

### *Updating Ratings*

Appropriateness ratings are updated as new data become available and resources permit. The 1989 coronary artery bypass graft ratings, for example, have been updated twice (in 1990 and 1994) (686).

[37] Most indications in the equivocal category are those for which there is agreement that the evidence of effectiveness is equivocal. But the category also includes indications for which there is disagreement among the panelists regarding appropriateness.

# Appendix D: Abbreviations and Glossary

D

## ABBREVIATIONS

| | |
|---|---|
| ABMT/HDC | autologous bone marrow transplant/high-dose chemotherapy |
| ACC | American College of Cardiology |
| ACIP | Advisory Committee on Immunization Practices (CDC) |
| ACOG | American College of Obstetricians and Gynecologists |
| ACP | American College of Physicians |
| AGA | American Gastroenterological Association |
| AHA | American Hospital Association |
| AHCPR | Agency for Health Care policy and Research |
| AIDS | acquired immunodeficiency syndrome |
| ALD | adenoleukodystrophy |
| ALLHAT | Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial |
| AMA | American Medical Association |
| AMCC | Academic Medical Center Consortium |
| AMI | acute myocardial infarction |
| APSAC | anisoylated plasminogen-activator complex |
| ASPN | Ambulatory Sentinel Practice Network |
| AUA | American Urological Association |
| BCBSA | Blue Cross and Blue Shield Association |
| BPH | benign prostatic hyperplasia |
| CABG | coronary artery bypass graft surgery |
| CCOP | Community Clinical Oncology Program (NCI) |
| CCU | coronary care unit |
| CDC | Centers for Disease Control and Prevention |
| CEA | cost-effectiveness analysis |
| CEAP | Clinical Efficacy Assessment Project (ACP) |
| CHAMPUS | Civilian Health and Medical program of the Uniformed Services |
| CHF | congestive heart failure |
| CHQC | Cleveland Health Quality Choice project |
| CPR | cardiopulmonary resuscitation |
| CQI | continuous quality improvement |
| CRC | colorectal cancer |
| CUA | cost-utility analysis |

| | |
|---|---|
| DATTA | Diagnostic and Therapeutic Technology Assessment program (AMA) |
| DES | diethylstilbestrol |
| DHHS | U.S. Department of Health and Human Services |
| ESRD | end-stage renal disease |
| EU | European Union |
| FDA | Food and Drug Administration |
| FOBT | fecal occult blood testing |
| GAO | U.S. General Accounting Office |
| GHPS | Group Health of Puget Sound |
| GISSI | The Gruppo Italiano per 10 Studio della Streptochinasi Nell' Infarto Miocardio |
| GUSTO | Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries |
| HCFA | Health Care Financing Administration |
| HCHP | Harvard Community Health Plan |
| HIE | RAND Health Insurance Experiment |
| HIV | human immunodeficiency virus |
| HMO | health maintenance organization |
| HPB | High Blood Pressure (educational program of NHLBI) |
| ICU | intensive care unit |
| IOM | Institute of Medicine |
| ISIS | International Study of Infarct Survival |
| JAMA | Journal of the American Medical Association |
| JCAHO | Joint Commission on Accreditation of Healthcare Organizations |
| MEDTEP | Medical Technology Effectiveness Program (AHCPR) |
| MMAF | Maine Medical Assessment Foundation |
| MMWR | Morbidity and Mortality Weekly Reports |
| MOS | Medical Outcomes Study |
| NCHCT | National Center for Health Care Technologies |
| NCHSR | National Center for Health Services Research |
| NCI | National Cancer Institute |
| NCQA | National Committee for Quality Assurance |
| NGT | Nominal Group Technique |
| NHLBI | National Heart, Lung, and Blood Institute |
| NICHD | National Institute for Child Health and Human Development |
| NIH | National Institutes of Health |
| NIMH | National Institute of Mental Health |
| NLM | National Library of Medicine |
| ODPHP | Office of Disease Prevention and Health Promotion |
| OHTA | Office of Health Technology Assessment (AHCPR) |
| OMAR | Office of Medical Applications of Research (NIH) |
| OTA | Office of Technology Assessment (U.S. Congress) |
| PDQ | Physician Data Query database |
| PHS | Public Health Service |
| PIVOT | Prostate Cancer Intervention Versus Observation Trial |
| PORT | Patient Outcomes Research Team |
| PSA | prostate-specific antigen |
| QALY | quality-adjusted life year |
| QWB | Quality of Well-Being scale |
| QOL | quality of life |
| R&D | research and development |
| RCT | randomized controlled trial |
| SEER | Surveillance, Epidemiology, and End Results database |
| SF-36 | RAND Short Form 36 |
| SIP | Sickness Impact Profile |
| TPA | tissue-type plasminogen activator |
| TURP | transurethral resection of the prostate |
| UR | utilization review |

| | |
|---|---|
| USPSTF | U.S. Preventive Services Task Force |
| VA | Department of Veterans Affairs |

## GLOSSARY

**AIDS (acquired immunodeficiency syndrome)**
A disease caused by the HIV virus and characterized by a deficiency of the immune system.

**Acute myocardial infarction**
A type of heart attack. Usually caused by insufficient blood supply to the heart, in myocardial infarction the sudden death of heart muscle is characterized by severe, unremitting chest pain, leading to arrhythmia (irregular heartbeat) and/or heart failure.

**Administrative databases**
Computerized databases that contain the administrative records of health insurers and health care providers (e.g., information from health insurance claims). These large databases contain information on medical, demographic, financial, and other characteristics of patients and their care that are useful for describing many aspects of health care.

**Algorithm**
A format for presenting a clinical practice guideline that consists of a structured flowchart of decision steps and preferred clinical management pathways. An algorithm prescribes what sequence of steps to take given particular circumstances or characteristics (e.g., a particular lab test result). Some algorithms also include designated points in the decisionmaking process where physicians and other caregivers need to discuss with patients or families their preferences for particular options.

**Appendectomy**
Surgical removal of the appendix.

**Appropriate**
Appropriate care is commonly defined as treatment that is considered to be effective and suitable in a specific clinical situation (i.e. for a particular patient) (compare with effectiveness.) Appropriateness research is a line of health services research that attempts to define the appropriate uses of particular medical technologies or practices.

**Benign prostatic hyperplasia**
A noncancerous enlargement of the prostate gland.

**Beta-blockers**
Drugs sometimes used to treat high blood pressure, irregular heartbeats, chest pain, stroke, and other cardiovascular disorders.

**Bias**
Systematic error introduced into a study. Bias can be introduced unconsciously into nonexperimental clinical studies by any characteristic or risk factor that systematically affects the results but is not actually due to the intervention being studied.

**Blinding**
In randomized controlled trials, keeping secret which treatment is assigned to participants. When only the patient is kept unaware of his or her treatment assignment, the study is single-blind. When the person administering the treatment (e.g. the physician) also is unaware, the study is double-blind. Additional layers of blinding can be added--e.g., when a third individual (usually the evaluator of outcomes, the individual analyzing data) also is unaware of treatment assignments.

**Carotid endarterectomy**
Surgical removal of the inner layer of the carotid artery (a major blood vessel leading to the head) when it is thickened and obstructed. Carotid endarterectomy is sometimes performed to lower the risk of a future stroke.

**Case-control study**
An observational epidemiological study that starts with the identification of a group of individuals with a disease (or other condition or outcome variable) of interest (cases), and a suitable group of persons without the disease. but who are otherwise similar to the cases (controls). The relationship of a risk factor or other attribute (e.g.. exposures to a chemical or physical agent, past receipt of a screening test) to the disease is evaluated by determining how frequently the risk factor is present in the cases and controls.

**Cholecystectomy**
Surgical removal of the gall bladder.

**Clinical practice guidelines**
The Institute of Medicine defines clinical practice guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." However, guidelines can also be developed with additional goals explicitly in mind, such as cost containment or reduction of defensive medicine. In this report, clinical practice guidelines that address medical technologies and practices, and that are created through a structured format of synthesis and analysis, are considered a special category of health technology assessments.

**Clinical-economic trial**
A *clinical trial (e.g.,* an RCT) that includes an economic data collection component, usually for use in a cost-effectiveness analysis.

**Clinical trial**
The systematic investigation of the effects of materials or methods (e.g., a medical technology) on humans in a clinical setting. Clinical trials can be either nonrandomized (e.g., a small trial to test a drug for major side effects) or randomized (see *randomized controlled trial).*

**Cohort study**
A study in which the experiences of a defined group (the cohort) are followed over time to link patient and/or clinical management characteristics with various health outcomes. A hypothetical cohort might be children living in homes with high levels of radon in 1980; a cohort study might examine how these children fare over time, and whether different subgroups of children fare differently.

**Colonoscopy**
A procedure that uses a long, flexible, fiberoptic instrument to visualize the large intestine (e.g., to detect signs of colon cancer). If abnormalities are seen, a biopsy (tissue sample) can be taken as part of the same procedure.

**Comorbidity**
Diseases or conditions present at the same time as the principal condition of a patient.

**Comparative effectiveness**
The effectiveness of two or more health care interventions relative to each other--+. g., which of two medical technologies leads to better health outcomes. See *effectiveness* and *effectiveness research.*

**Congestive heart failure**
A chronic medical condition usually caused by a heart disorder and retention of salt and water by the kidneys. It is characterized by shortness of breath, prolonged circulation time, and edema (swelling) of the extremities.

**Consensus**
In the context of the group process used to derive a *clinical practice guideline,* consensus does not necessarily mean unanimous agreement. Often, consensus merely means that most members of the group agree. A few guideline development efforts have developed definitions of consensus that specify in detail the level of agreement that group members must have in order for consensus to be achieved.

**Continuous quality improvement (CQI)**
An array of formal quality improvement techniques based on the collection and analysis of data generated in the course of current clinical practice in a defined clinical setting in order to identify and solve problems in the system.

**Contrast agent**
A substance that is used to improve the visibility of structures during radiologic imaging+. g., angiography, intravenous urography, or computerized tomography scans. A low-osmolality contrast agent has an osmolality (i.e., concentration of dissolved particles in solution) that is closer to the osmolality of body fluids than the osmolality of traditional contrast agents.

**Coronary angiography**
The radiological visualization of the coronary arteries (the blood vessels that feed the heart muscle), after injection of a contrast agent.

**Coronary artery bypass graft (CABG) surgery**
**A** surgical procedure to treat coronary artery disease in which a vein or an artery is used to bypass a

constricted portion of one or more coronary arteries.

**Cost-benefit analysis**
An analytic technique that enumerates and compares the net costs of an intervention (e.g. a medical technology or a public health program) with the net cost savings that arise as a consequence of applying that intervention. Results are expressed entirely in monetary units.

**Cost-effectiveness analysis**
An analytic technique that calculates the cost per specified health effect of a technology or program (e.g., cost per lives saved, or cost per cases of cancer avoided) and compares this cost-effectiveness ratio with ratios from other interventions. Effects of all interventions are expressed in similar values. In contrast to *cost-benefit analysis,* health outcomes are not translated into monetary values.

**Cost-utility analysis**
A form of cost-effectiveness analysis in which values are assigned to different kinds of health outcomes, reflecting the relative importance of the different outcomes to people, and results are expressed in units such as cost per *quality-adjusted life year*. Expressing results in this way facilitates comparisons across health care interventions with very different effects (e.g., saving lives vs. reducing disability).

**Cross-sectional study**
A study that examines the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at one particular time (compare with *cohort study).*

**Decision analysis**
A technique to aid decisionmaking under conditions of uncertainty. A decision analysis involves the systematic, schematic representation and examination of all of the relevant information for a decision, the points at which decisions or uncertain events occur, and the relative preferences the decisionmaker would have for the array of various possible outcomes for a decision.

**Delphi technique**
A formal process for eliciting judgments from a group that involves a set of iterative cycles in which individuals in a group provide their ratings or opinions. A summary of the responses is made and given back to the individuals, and the individuals are then asked to re-evaluate their previous ratings. The Delphi technique does not require that the individuals in the group actually meet in person (although they may).

**Detailing**
The use of personal contacts between a designated liaison individual and a physician as a mechanism for encouraging that physician (or other health care professional) to alter his or her behavior. The detail person might be, for example, a drug compan y representative trying to convince a physician to prescribe a particular brand of drug more often. or another clinician who is trying to convince the physician to follow a particular clinical practice guideline.

**Diagnosis-related group (DRG)**
Entries in a taxonomy of types of hospitalizations based on groupings of diagnostic categories drawn from the *International Classification of Diseases* and modified by the presence of a surgical procedure, patient age, presence or absence of significant comorbidities or complications, and other relevant criteria. DRGs have been inundated for use in establishing payment amounts for individual admissions under Medicare prospective hospital payment system as required by the Social Security Amendments of 1983 (Public Law 98-21 ).

**Discounting**
A procedure used in economic analysis (e.g. cost-effectiveness analysis) to express as ‘present values” those costs and benefits that will occur in future years. Discounting is based on two premises: 1 ) individuals prefer to receive benefits today rather than in the future; and 2) resources invested today in alternative programs could earn a return over time.

**Effectiveness**
The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under average or actual conditions of use. Compare with *efficacy.*

**Effectiveness research**
The category of research efforts aimed at broadly identifying effective technologies and practices, and developing and refining methods to support the identification of effective care. Effectiveness research includes experimental research that is aimed at gaining information on effectiveness of an intervention for a broad population in community health care settings.

**Efficacy**
The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use. Efficacy is generally evaluated in controlled trials of an experimental therapy and a control condition. Compare with *effectiveness.*

**End-stage renal disease (ESRD)**
Chronic renal failure that occurs when an individual irreversibly loses a sufficient amount of kidney function so that life cannot be sustained without treatment (e.g., hemodialysis, kidney transplant surgery, or continuous ambulatory peritoneal dialysis).

**Endoscopy**
Inspection of an internal cavity using a fiberoptic imaging device.

**Epidemiology**
The scientific study of the distribution and occurrence of human diseases and health conditions, and their determinants.

**Experimental study**
**A study** in which the experiences of the groups being studied are intentionally influenced by the researcher. The *randomized controlled trial is* a form of experimental study common in the medical sciences.

**Firms trial**
**A** form of *randomized controlled trial* in which patients are randomized among entire clinics or other institutional settings (i.e., the patients are randomized to experimental and control clinics, rather than only to experimental and control treatments within a single clinic).

**Formulary**
A list of selected pharmaceuticals and their appropriate dosages judged to be the most useful or cost-effective for patient care from which physicians are required or encouraged to prescribe. A formulary may also be a list of drugs that may not be prescribed without special appeals.

**Generalizability**
**See** *validity (external validity).*

**Geographic variation**
In the context of health services research, variations across geographic areas (e.g., states) in the rates with which particular medical procedures are performed, the costs of medical care, or other variables of interest.

**Generic measure of health-related quality of life**
A measure of *health-related quality of life* that is designed to be broadly applicable across diseases and health conditions, populations, and health care interventions and that summarizes individuals' health across the various aspects of health-related quality of life. In contrast, a disease-specific instrument measures only those aspects applicable to a single disease or condition.

**Health maintenance organization (HMO)**
A health care organization that, in return for prospective per capita (cavitation) payments, acts as both insurer and provider of comprehensive but specified health care services. A defined set of physicians (and. often, other health care providers such as physician assistants and nurse midwives) provide services to a voluntarily enrolled population. Prepaid group practices and individual practice associations, as well as staff models, are types of HMOS .

**Health services research**
The integration of epidemiological, sociological, economic, and other analytic sciences in the study of health services. Health services research is concerned with relationships between need, demand. supply, use, cost, and outcome of health services.

**Health technologies**
Drugs, devices. procedures, and the organizational and support systems within which health care is delivered.

**Health technology assessment**
**A structured** analysis of a health technology, a set of related technologies, or a technology-related issue that is performed for the purpose of providing input to a policy decision.

**Hypertension**
**High blood** pressure.

**Hysterectomy**
Surgical removal of the uterus.

**Immunotherapy**
The treatment of disease by the administration to the patient of an antibody raised in another individual or another species (passive immunotherapy) or by immunizing the patient with antigens appropriate to the disease (active immunotherapy).

**Longitudinal study**
**See** *cohort study.*

**Laparoscopic cholecystectomy**
The surgical removal of the gall bladder and its contents through a small incision, with the aid of an imaging device.

**Lidocaine**
**A drug** that acts as a local anesthetic but that, injected into the bloodstream, is also sometimes used to treat abnormal heartbeat and *acute myocardial infarction.*

**Managed care**
**A general term** applied to a range of initiatives from organized health care delivery systems (e.g., HMOs) to features of health care plans (e.g., preadmission certification programs. *utilization? review* programs) that attempt to control or coordinate enrollees' use of (and thus control the cost of) services.

**Medical technology**
**See *health technology.***

**Medicaid**
A joint federal/state program intended to improve health care and heidth-related services for low-income individuals. Medicaid regulations are established by each state within federal guidelines. and the eligibility requirements and services covered vary significantly among the states. In general. Medicaid pays for medical. nursing home, and home health care for individuals who meet the eligibility requirements for those services. Financial eligibility for Medicaid is determined by a means test, in which a ceiling is placed on the maximum income and assets an individual may have in order to qualify for assistance.

**Medicare**
A nationwide, federally administered health in surance program authorized by Title XVIII of the Social Security Act of 1965 to cover the cost of hospitalization, medical care, and some related services for eligible persons over the age of 65. persons receiving Social Security Disability Insurance payments for two years, and persons with end-stage renal disease. Medicare consists of two separate but coordinated programs-hospital insurance (Part A) and supplementary medical insurance (Part B). Health insurance protection is available to insured pm-sons without regard to income.

**Meta-analysis**
A systematic, quantitative review of a subject. Three major features distinguish this method from a traditional narrative review: the formal and comprehensive search for relevant data: the explicit, objective criteria for selecting studies to be included; and the quantitative statistical analysis of the studies results. See also *systematic review.*

**Morbidity**
The condition of being ill or othcrwise afflicted with an unhealed condition.

**Multiple sclerosis**
A progressive, chronic disease in which nerve fibers of the brain and spinal cord lose their myelin cover and do not function properly, often leading to difficulty walking and other disabilities.

**Nominal Group Technique (NGT)**
A formal technique for managing interactions among individuals in a group, The NGT splits problem solving into two phases, an idea-generating phase and a decisionmaking phase. In the first phase, each member of the group individually makes a list of ideas for group consideration. All individually generated ideas are then recorded on a flip chart for the group and are openly discussed, In the second decisionmaking phase of the process, individuals vote on priority ideas and a group decision is mathematically derived through rank-ordering or rating.

**Observational study**
A study in which the actual experiences of the groups being compared are simply observed, often retrospectively (i.e., after the event of interest, such as having surgery or exposure to a toxic substance, has occurred).

**Off-label use**
The prescription or use of ethical pharmaceuticals for indications other than those specified in FDA-approved labeling of the dreg.

**Opinion leader**
An influential clinical peer or respected expert in the field.

**Outcome**
Any result that stems from exposure to a causal factor, or from preventive or therapeutic interventions.

**Outcomes research**
**A** term originally used to describe a particular line of health services research that focused on identifying variations in medical procedures and associated health outcomes. The term has since been applied to a wide variety of vaguely associated activities and no longer has a clearly identifiable meaning. See effectiveness *research.*

**Patient Outcomes Research Teams (PORTS)**
Multidisciplinary research teams, funded by the Agency for Health Care Policy and Research, that investigate particular health care conditions, with attention to variations in how those conditions are managed and the effectiveness and costs of the technologies used to treat them.

**Patient preferences**
**A patient's judgments** of the desirability of a particular set of health outcomes or situations (see also *utility).*

**Payer**
An entity that pays for health care services (e.g., individuals, health insurers, government programs). Third-party payers are payers other than the individuals receiving the services, usually health insurers.

**Performance indicators**
Also sometimes called report cards or scorecards, performance indicators are measures that can be used to rate providers, insurers, or health care plans according to their performance along several criteria. Common indicators include mortality rates, costs, rates of specific procedures, or rates of hospitalization for preventable diseases.

Perinatal
Pertaining to or occurring in the period shortly before and after birth; variously defined as beginning with the completion of the to the twenty-eighth week of gestation and ending seven to 28 days after birth.

**Placebo**
A drug or procedure with no intrinsic therapeutic value. In a *randomized controlled trial,* a placebo is given to patients in control groups as a means to blind investigators and patients as to whether an individual is receiving the experimental or the control treatment. See *blinding.*

**Practice profiling**
Expressing a provider's pattern of practice as a rate—some measure of use (costs or services) or outcome (functional status, morbidity or mortality) aggregated over time for a defined **population of patients** under the provider-s care-for the pur-

pose of comparison with other providers' practice patterns.

**Prospective payment**
Payment for medical care on the basis of rates set in advance of the time period in which they apply. The unit of payment may vary from individual medical services to broader categories, such as hospital case, episode of illness, or person (capita-(ion). Medicare's DRG payment system for inpatient hospital services is a particular form of prospective payment.

**Prostatectomy**
**Surgical** removal of the prostate gland.

**Protocol**
The plan or outline of a scientific experiment, treatment, or study.

**Provider**
A person or organization that provides health care services (e.g., physician, optometrist, hospital, home health agency).

Publication bias
Occurs when the published studies are not representative of the results of all studies that have been conducted on the research question. Publication bias reflects the preference for publishing studies that have statistically significant findings or that support popular ideas.

**Quality of care**
Evaluation of the performance of medical providers according to the degree to which the process of care increases the probability of outcomes desired by patients and reduces the probability of undesired outcomes, given the state of medical knowledge. Which elements of patient outcomes predominate depends on the patient condition.

**Quality of life**
In the context of *effectiveness research* and *cost-effectiveness analysis,* health-related quality of life is "the value assigned to duration of life as modified by the impairments, functional states. perceptions and social opportunities that are influenced by disease, injury, treatment, or policy" (591 ).

**Quality-adjusted life year (QALY)**
Years of life saved by a technology or service, adjusted according to the quality of those lives (as determined by some valuation process; see also *utilities).* The QALY is the most commonly' used unit to express the results of cost-utility analyses.

**Randomized controlled trial**
An *experimental study* designed to test the safety, efficacy, or effectiveness of a health care interven - tion, in which people are randomly allocated to experimental or control groups, and the outcomes are compared. The experimental group or groups receive the intervention of interest. while the control group receives a *placebo* or usual care.

**Reliability**
The reproducibility of a measure. A measure is reliable if it yields similar results each time it is used on similar samples, or if its components yield similar results for the same or similar samples (compare *validity).*

**Scoliosis**
An abnormal lateral (sideward) curve of the spine.

**Sensitivity analysis**
An analysis of the effect of changes in key assumptions or uncertainties on the findings and outcomes of a scientifically conducted study.

**Sickle-cell disease**
A hereditary disease in which the body's red blood cells are sickle shaped and fragile and can become plugged in small blood vessels.

**Sigmoidoscopy**
An *endoscopic* procedure for visualizing the sigmoid colon (the lower portion of the largc intestine) (see also *colonoscopy).*

**Standard of care**
A legal standard defined as the level of care provided by the majority of physicians in a particular clinical situation. In a malpractice action, a physician's actions are judged against the prevailing standard of care.

**Statistical power**
The probability of detecting a statistically significant difference between the groups being

compared in a randomized experiment when a real difference does exist.

**Statistical significance**
**A statistically** significant finding is one that is unlikely to have occurred solely as a result of change. A finding is typically considered to be statistically significant if the probability that it occurred by chance alone is no greater than five out of 100-i. e., a p value of 0.05 or less (although the usefulness of having an exact cutoff level is widely debated).

**Stroke**
**A** condition in which a hemorrhage (internal bleeding) or blockage of blood vessels leading to the brain results in a lack of oxygen in the brain tissue and a sudden loss of functional ability.

**Systematic review**
A highly structured review that synthesizes the results of previously conducted studies on a particular health topic. A *meft-analysis* is one type of systematic review.

**Thrombolytic drugs**
Drugs used to promote the dissolution of blood clots within the circulatory system.

**Thrombus**
**A** blood clot attached to the internal wall of a blood vessel.

**Toxicity**
The quality of being poisonous or the degree to which a substance is poisonous. Referring to medical treatments, the degree to which they produce unwanted, adverse effects.

**Transurethral prostatectomy**
*See transurethral resection of the prostate (TURP).*

**Transurethral resection of the prostate (TURP)**
**A** surgical technique involving the insertion of an instrument through the urethra (urinary conduit) in order to remove tissue from the prostate gland.

**Treatment effect**
The outcome measure (e.g.. the rate of death or the proportion of subjects who died) in the treated group compared with that in the control group.

**Utility**
In decision analysis and economics, a concept referring to the desirability of, or preference for, a particular health outcome or health state. Utilities are quantitative—i.e., they describe not only whether outcome A is preferred to outcome B, but how much is preferred.

**Utilization review**
The review of services delivered by a health care provider or supplier to determine whether those services were medically necessary.

**Validity**
Broadly defined, validity is the extent to which an observed situation reflects the true situation. *Internal* validity is a measure of the extent to which study results reflect the true relationship of an intervention to the outcome of interest in the study subjects. *External* validity is the extent to which the results of a study may be generalized beyond the subjects of the study to other settings, providers, procedures, diagnostics, etc. (compare *reliability).*

# References


1. Aaron, H. J., *Serious and Unstable Condition: Financing America's Health Care* (Washington, DC: Brookings Institution, 1991).
2. Abrams, P. H., and Griffiths, D. J., "The Assessment of Prostatic Obstruction from Urodynamics Measurements and from Residue Urine," British Journal of Urology51: 129, 1979.
3. Abt Associates Inc., Bethesda, MD, "Estimating Increases in Medicare Costs Due to Scientific and Technological Advances," unpublished contractor report prepared for the Physician Payment Review Commission, contract no. T-99388649, Washington, DC, January 1994.
4. Achusim, L. E., Weller, T. W., Somani, S. M., et al., "Comparison of Automated and Manual Methods of Syringe Fill ing," *American Journal of Hospital Pharmacy* 47:2492-2495,1990.
5. Adams, M. E., McCall, N. T., Cray, D. T., et al., "Economic Analysis in Randomized Control Trials," *Medical Care* 30(3):231-*243,* 1992.
6. "Advances in Health Status Assessment: Proceedings of a Conference," *Medical Care, 30(5)* (suppl.): May 1992.
7. "AHCPR," *Health News Daily,* p. 7, June 9, 1994.

7a. "AHCPR Seeks Input on Guidelines Topics — Reauthorization Calls for at Least Three Prevention Guidelines by 1996," *Abstracts of Clinical Care Guidelines 5(8):* 18-19, 1993.

8. Alliance for Health Reform, briefing on "A Consensus Approach to Ensuring Quality Care Under Health Reform," presentations by R. H., Brook, P.D. Cleary, D.M. Eddy, et al., Washington, DC, Apr. 26, 1994.
9. Altman, L. K., "Surgery Found to Help Emphysema Patients," *New York Times,* p. A-18, Apr. 28, 1994.
10. "AMA Identifies First Guidelines Found 'in Conformance' with Attributes for Guideline Development," *Abstracts of Clinical Care Guidelines 5(* 10): 19-20, December 1993.
11. Ambulatory Sentinel Practice Network, Inc. (ASPN), "ASPN: A Practice-Based Research Network of North American Primary Care Clinicians in Association with the American Academy of Family Physicians," Denver, CO, undated.
12. Ambulatory Sentinel Practice Network, Inc. (ASPN), "Background of ASPN," unpublished document, Denver, CO, undated.
13. American College of Physicians, "Guidelines for CEAP Authors," unpublished document, Philadelphia, PA, August 1988.
14. American College of Physicians, "Practice Strategies for Elective Red Blood Cell Transfusion," *Annals of Internal Medicine*, 11 6(5): 403-406, 1992.
15. American College of Physicians, "Automated Ambulatory Blood Pressure and Self-Measured Blood Pressure Monitoring Devices: Their Role in the Diagnosis and Management of Hypertension," *Annals of Internal Medicine* 118(1 1 ):889, 1993.
16. American College of Physicians, *Clinical Practice Guidelines Approved by the American College of Physicians, 1992* Edition (Philadelphia, PA: 1993).
17. American College of Physicians, "Guidelines for the Treatment of Gallstones." *Annals of Internal Medicine* 11 *9(7,* part 1):620-622, 1993.
18. American College of Physicians, American Diabetes Association, American Academy of Ophthalmology, "Screening Guidelines for Diabetic Retinopathy," *Annals of Internal Medicine* 11 6(8):683-685, 1992.

19. American Medical Association, Diagnostic and Therapeutic Technology Assessment, "Chronological Index of DATTA Evaluations," Chicago, IL, undated.
20. American Medical Association, Diagnostic and Therapeutic Technology Assessment, "DATTA Procedures," Chicago, IL, April 1992.
21. American Medical Association, Office of Quality Assurance and Medical Review, *Directory of Practice Parameters. Titles, Sources, and Updates* (Chicago, IL: 1993).
22. American Medical Association. Specialty Society Practice Parameters Partnership, "Attributes to Guide the Development of Practice Parameters," *The Medical Outcomes and Guidelines Sourcebook*, S. Vibbert and J. Reichard (eds.) (Washington, DC: Faulkner and Gray, 1993).
23. Andersen, J. T., Nordling, J., Walter, S., "Prostatism I. The Correlation Between Symptoms, Cytometric, and Urodynamic Findings," *Scandinavian Journal of Urology and Nephrology* 13:229, 1979.
24. Anderson, G. F., Steinberg, E. P., Whittle, J., et al., "Development of Clinical and Economic Prognoses from Medicare Claims Data," *Journal of the American Medical Association* 263:967-972, *1990.*
*25.* Anderson, G. F., Hall, M. A., and Steinberg, E. P., "Medical Technology Assessment and Practice Guidelines: Their Day in Court," *American Journal of Public Health* 83(11): 1635-1639, 1993.
26. Anderson, G. F., and Steinberg, E. P., "Hospital Readmissions in the Medicare Population." *New England Journal of Medicine* 311:1349-*1353, 1984.*
*27.* Antman, E., Lau, J., Kupelnick, B., et al., "A Comparison of the Results of Meta-Analyses of Randomized Controlled Trials and Recommendations of Clinical Experts," *Journal of the American Medical Association 268(2):240-248, 1992.*
*28.* Aspirin in Myocardial Infarction Study Group, "A Randomized Controlled Trial of Aspirin in Persons Recovering from Myocardial Infarction," *Journal of the American Medical Association* 243:661-*669, 1980.*
*29.* Auborg, P., Adamsbaum, C., Lavallard-Rousseau, M. C., et al., "A Two-Year Trial of Oleic and Erucic Acids ('Lorenzo's Oil') as Treatment for Adrenomyeloneuropathy," *New England Journal of Medicine 329(11):*745-52, 1993.
30. Audet, A. M., Greenfield, S., and Field, M., "Medical Practice Guidelines: Current Activities and Future Directions," *Annals (///' Internal Medicine* 113:709-714, 1990.
31. Avorn, J. L., and Soumerai, S. B., "Improving Drug Therapy Decisions Through Educational Outreach: A Randomized Controlled Trial of Academically-Based 'Detailing'," *New England Journal of Medicine* 308: *1457-1463, 1983.*
*32.* Avorn, J. L., Soumerai, S. B., Everitt, D. E., et al., "A Randomized Trial of a Program to Reduce the Use of Psychoactive Drugs in Nursing Homes," *New England Journal of Medicine* 327: 168-173, 1992.
33. Ayanian, J. Z., and Epstein, A. M., "Differences in the Use of Procedures Between Men and Women Hospitalized for Coronary Artery Disease," *New England Journal of Medicine* 325(4):221-225, 1991.
34. Ayanian, J. Z., Udvarhelyi, 1. S., and Gatsonis, G. A., "Racial Differences in the Use of Revascularization Procedures After Coronary Angiography," *Journal of the American Medical Association* 269:2642-2646, 1993.
*35.* Baker, M. S., Kessler, L. G., Urban, N., et al., "Estimating the Treatment Costs of Breast and Lung Cancer," *Medical Care* 29:40-49, 1991.
36. Bakketeig, L. S., "Methodological Problems and Possible Endpoints in the Evaluation of Antenatal Care," *International Journal of Technology Assessment in Health Care* 8(suppl. 1):33-39, *1992.*
*37.* Baines, C. J., "The Canadian National Breast Screening Study: a Perspective in Criticisms," *Annals of Internal Medicine* 120(4):326-334, February 1994.
38. Barnes, B., '. Discarded Operations: Surgical Innovation by Trial and Error," *Costs, Risks, and Benefits of Surgery*, J.P. Bunker, B.A. Barnes. and F. Mosteller (eds.) (New York, NY: Oxford University Press, 1977).

39. Barnes, B. A., O'Brien, E., Comstock, C., et al., "Report on Variation in Rates of Utilization of Surgical Services in the Commonwealth of Massachusetts," *Journal of the American Medical Association* 254(3):37 *1-375,* 1985.

40. Barnett, G. O., Winickoff, R. N., Morgan, M. M., et al., "A Computer-Based Monitoring System for Follow-Up of Elevated Blood Pressure," *Medical Care 21* :400-409, *1983.*

*41.* Barry, M. J., Director, Health Services Research Program, Massachusetts General Hospital, Boston, MA, personal communication, May 24, 1994.

42. Barry, M. J., Fowler, F. J., O'Leary, M. P., et al., "Correlation of the American Urological Association Symptom Index with Self-Administered Versions of the Madsen- Iversen, Boyarsky and Maine Medical Assessment Program Symptom Indexes," *Journal of Urology 148: 1558-1563, 1992.*

*43.* Barry, M. J., Cockett, A. T. K., Holtgrewe, H. L., et al., "Relationship of Symptoms of Prostatism to Commonly-Used Physiological and Anatomical Measures of the Severity of Benign Prostatic Hyperplasia," *Journal of Urology 150(2* pt. 1 ):35 1-358, 1993.

44. Battista, R. N., and Fletcher, S. W., "Making Recommendations on Preventive Practices: Methodological Issues," *American Journal of Preventive Medicine 4(4)( suppl):53-67, 1988.*

*45.* Baumann, R., Magos, A. L., and Turnbull, S. A., "Prospective Comparison of Videopelviscopy with Laparotomy for Ectopic Pregnancy," *British Journal of Obstetrics and Gynecology* 98:765-771, *1991.*

*46,* Begg, C. B., and Berlin, J. A., "Publication Bias and Dissemination of Cl inical Research," *Journal of the National Cancer Institute 81 (2): 107-115, 1989.*

*47.* Bennett, C. L., Garfinkle, M. S., Greenfield, S, et al., "The Relation Between Hospital Experience and In-Hospital Mortality for Patients With AIDS-Related PCP," *Journal of the American Medical Association 261* (20):2975-2979, *1989.*

*48.* Bergman, R., "Making the Grade," *Hospitals* and *Health Networks* 68(1):34-36, Jan. 5, 1994.

49. Bergner, M, Bobbitt, R. A., Carter, W. B., et al., "The Sickness Impact Profile: Development and Final Revision of a Health Status Measure," *Medical Care* 19:787-805, 1981.

50. Bergner, M., Barry, M. J., Bowman, M. A., et al., "Where Do We Go From Here'? Opportunities for Applying Health Status Assessment Measures in Clinical Settings,*' Medical *Care* 30(5) (suppl. ): MS219-MS230, 1992.

51. Berlin, J. A., and Colditz, G. A., "A Meta-Analysis of Physical Activity in the Prevention of Coronary Heart Disease," *American Journal of Epidemiology* 132(4):61*2-628, 1990.*

*52.* Bernstein, J., Chief, Program Planning, Research, Development, and Evaluation Branch, Office of Program Development, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, January 1994.

53. Bernstein, J., Chief, Program Planning, Research, Development, and Evaluation Branch, Office of Program Development, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, letter to the Office of Technology Assessment, U.S. Congress, Washington DC, Apr. 22, 1994.

54. Bernstein, S. J., Hilborne, L. H., Leape, L. L., et al., "The Appropriateness of Use of Coronary Angiography in New York State," *Journal of the American Medical Association 269(6): 766-769, 1993.*

*55.* Bernstein, S. J., McGlynn, E. A., Siu, A. L., et al., "The Appropriateness of Hysterectomy: A Comparison of Care in Seven Health Plans," *Journal of the American Medical Association 269(*18):2398-2402, *1993.*

*56.* Bernstein Research, "The Future of Health-Care Delivery in America," data on pharmacy services provided by Ken Abromowitz, New York, NY, June 26, 1993.

57. Berwick, D. M., "Continuous Improvement as an Ideal in Health Care," *New England Journal of Medicine 320(1* ):53-56, *1989.*

*58.* Berwick, D. M., and Coltin, K. L., "Feedback Reduces Test Use in a Health Maintenance Organization," *Journal of the American Medical Association 255:* 1450-1454, 1986.

59. Bess, F., and Paradise, J. L., "Universal Screening for Infant Hearing Impairment: Not Simple,

Not Risk Free, Not Necessarily Beneficial, and Not Presently Justified,"' *Pediatrics* 93(2):330-334, February 1994.

60. Bezerra, J. A., Stathos, T. H., Duncan, B., et al., "Treatment of Infants With Acute Diarrhea: What's Recommended and What's Practiced," *Pediatrics 90: 1-4, 1992.*

*61.* Bibbo, M., Gill, W. B., Azizi, F., et al., "Follow-Up Study of Male and Female Offspring of DES-Exposed Mothers," *obstetrics and Gynecology* 49: 1-8, 1977.

62. Bibbo, M., Haenszel, W. M., Wied, G, L., et al., "A Twenty-Five Year Follow-Up Study of Women Exposed to Diethylstilboestrol During Pregnancy," *New* England *Journal of Medicine* 298:763-767, *1978.*

*63.* Black, N., Petticrew, M., Ginzler, M., et al., "DO Doctors and Patients Disagree? Views of the Outcome of Transurethral Resection of the Prostate," *International Journal of Technology Assessment in Health Cure* 7(4):533-544, 1991.

64. Blakeslee, S., "Program to Cut Risks of Diabetes Surprisingly Fails to Lure Patients," *New York Times,* pp. Al, B8, Feb. 28, 1994.

65. Bloom, B., and Fendrick, M., "Policy Effects of Minimumally Invasive Surgery," abstract presented at the International Society for Technology Assessment in Health Care meeting, Solento, Italy, May 23, 1993.

66. Blue Cross and Blue Shield Association, "The Blue Cross and Blue Shield Demonstration Project: Statement of Purpose," unpublished document, Chicago, IL, April 1991.

67. Blue Cross and Blue Shield Association, Health Benefits Management Division, Technology Management Department, *Medical Necessity Program Manual* (Chicago, IL: 1987).

68. Blumenthal, D., "Federal Policy Toward Health Care Technology: The Case of the National Center," *Milbank Quarterly 61* :584-613, 1983.

69. Boden, W. E., "Meta-Analysis in Clinical Trials Reporting: Has a Too] Become a Weapon'?" (editorial), *American Journal of Cardiology* 69(6):681-686, *1992.*

*70.* Boettcher, W. G., "Total Hip Arthroplasties in the Elderly: Morbidity, Mortality, and Cost Effectiveness," *Clinical Orthopedics and Related Research* 274:30-34, *1992.*

71. Boissel, J. P., "'impact of Randomized Clinical Trials on Medical Practices," *Controlled Clinical Trials 10:* 120S- 134S, 1989.

72. Borbas, C., McLaughlin, D., and Schultz, A., "Minnesota Clinical Comparison and Assessment Program: Bridging the Gap Between Clinical Guidelines and Clinical Practice," *Bridging the Gap Between Theory and Practice Guidelines* (Chicago, IL: Hospital Research and Educational Trust, 1993).

73. Borbas, C., Norling, R., McLaughlin, D. B., et al., "Frequently Asked Questions About the Minnesota Clinical, Comparison & Assessment Program (MCCAP): How Is It Used and Does It Work'?," MCCAP, Minnesota, MI, unpublished manuscript, May 1993.

74. Borzo, G., "Illinois Blues: If You Sign a Contract, You Follow Guidelines," *American Medical News* 36(45):8, Dec. 6, 1993.

75. Bowling, A., *Measuring Health: A Review of Quality of Life Measurement Scales* (Philadelphia, PA: Open University Press, 1991 ).

75a. Boyd, N. F., Sutherland, H. J., Heasman, K. Z., et al., "Whose Utilities for Decision Analysis?" *Medical Decision Making 10(1 ):58-67, 1990.*

*76.* Boyle, J. F., Vice Chairman, Board of Trustees of the American Medical Association, Ch icago, IL, "State merit," *Health Care Technology Reauthorization*, hearing before the Subcommittee on Health and the Environment, Committee on Energy and Commerce, House of Representatives, U.S. Congress, Mar. 20, 1981, Serial No. 97-27 (Washington, DC: U.S. Government Printing Office, 198 l).

77. Brackbill, Y., and Berendes, H. W., "Dangers of Diethylstilboestrol: Review of a 1953 Paper," (letter) *Lancet* 2(8088):520, September 1978.

78. Bray, E., Program Analyst, Office of Medical Applications of Research, Office of the Director, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, personal communication, May 20, 1993.

79. Brechner, R. J., Cowie, C. C., Howie, L. J., et al., [b]* Ophthalmic Examination Among Adults With Diagnosed Diabetes Mellitus," *Journal of the* American *Medical Association 270(1 4):* 1714-*1718, 1993.*

80. Brittan, D., 'phenomena: Roving Glands," *Technology Review* 96(8 ):78, 1993.
81. Brook, R., RAND Corp., Santa Monica, CA, personal communication, April 1994.
82. Brook, R. H., Chassin, M. R., Fink, A., et al., "A Method for the Assessment of the Appropriateness of Medical Technologies," *International Journal of Technology Assessment in Health Care* 2(1 ):53-63,1986.
83. Brook, R. H., Kamberg, C. J.. Mayer-Oakes, A., et al., "Appropriateness of Acute Medical Care for the Elderly: An Analysis of the Literature," *Health Policy* 14:225-242, 1990.
84. Brook, R. H., Park, R. E., Chassin, M. R.. et al., "Predicting the Appropriate Use of Carotid Endarterectomy, Upper Gastrointestinal Endoscopy, and Coronary Angiography," *New England Journal of Medicine* 323( 17):1 173-1177, 1990.
85. Brook, R. H., Park, R. E., Winslow, C. M., et al., "Diagnosis and Treatment of Coronary Disease: Comparison of Doctors' Attitudes in the USA and the UK," *Lancet* 1(8588):750-753, 1988.
86. Brook, R. H., Ware, Jr., J., Rogers, W., et al., "The Effect of Coinsurance on the Health of Adults," R-3055-HHS (Santa Monica, CA: RAND Corp., December 1984).
87. Brooks, S. L., "Survey of Compliance With American Heart Association Guidelines for the Prevention of Bacterial Endocarditis," *Journal of the American Dental Association* 101:41-43, 1980.
88. Brown, B. W., "The Randomized Clinical Trial," *Statistics in Medicine* 3:307-311, 1984.
89. Brown, M. L., and Fintor, L.. "Cost-Effectiveness of Breast Cancer Screening: Preliminary Results of a Systematic Review of the Literature," *Breast Cancer Research and Treatment* 25(2): 113-118, 1993.
90. Brown, S. A., 'measurement of Quality of Primary Studies for Meta-Analysis," *Nursing Research* 40(6):352-355. 1991.
91. Bulpitt, C., "Meta-Analysis," *Lancet* 2(8602): 93-94, 1988.
92. Buring, J. E., Jonas, M. A., Hennekens, C. H., Division of Preventive Medicine, Brigham and Women's Hospital and the Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA, "Large and Simple Randomized Trials," paper prepared under contract for the Office of Technology Assessment, U.S. Congress, Washington, DC. forthcoming 1994.
93. Burns, L. R., Denton, M., Goldfe in, S.. et al., "The Use of Continuous Quality Improvement Methods in the Development and Dissemination of Medical practice Guidelines," *Quality Review Bulletin* 18( 12):434-439, 1992.
94. Byar, D. P., "Why Data Bases Should Not Replace Randomized Clinical Trials," *Biometrics* 36:337-342, June1 980.
95. Byar, D. P., "The Use of Data Bases and Historical Controls in Treatment Comparisons," *Recent Results in Cancer Research* 111:95-98, 1988.
96. Byar, D. P., "Design Considerations for AIDS Trials." *Journal of Acquired Immune Deficiency Syndrome* 3(suppl. 2): S16-S19,1990.
97. Byar, D. P., "Problems With Using Observational Databases to Compare Treatments" *Statistics in Medicine* 10(4):663-666, 1991.
98. Byar, D. P., Schoenfeld, D. A., Green. S. B.. et al.. "Design Considerations for AIDS Trials," *New England Journal of Medicine* 323: 1343-1348, 1990.
99. California Public Employees' Retirement System, Health Benefits Advisory Council, *Designing a Fair and Reasonable Basic Benefit Plan Using Clinical Guidelines*, summary of a conference sponsored by Cal PERS, (Sacramento. CA: Information and Program Development Division, CalPERS, 1992).
100. Canadian Task Force on the Periodic Health Examination, "The Periodic Health Examination," *Canadian Medical Association Journal* 121:1 193-1254,1979.
101. Canner, J. K., Javitt, J. C., and McBean, A. M., "National Outcomes of Cataract Extraction 111. Cornea] Edema and Transplant Following Inpatient Surgery," *Archives of Ophthalmology* 110:1137-1142, 1992.
102. Caplow, T.. "Market Attitudes: A Research Report From the Medical Field," *Harvard Business Review* 30: 105-112,1954.
103. Carlson, K. J., Miller, B. A., Fowler, F.J. Jr., "The Maine Women's Health Study: I Outcomes of Hysterectomy," *American Journal of*

*Obstetrics and Gynecology* 83(4):556-565, April 1994.

104. Cargill, V., Conti, M., Neuhauser, D., et al., "* Improving the Effectiveness of Screening Colorectal Cancer by Involving Nurse Clinicians," *Medical Care* 29( 1): 1-5, 1991.
105. Cebul, R., ● 'Randomized, Controlled Trials Using the Metro Firm System," *Medical Care 29(7):JS9-18, 1991.*
*106,* Chalmers, I., "Evaluating the Effects of Care During Pregnancy and Childbirth," *Effective Care in Pregnancy and Childbirth, Volume I: Pregnancy, Parts Z-W, 1.* Chalmers, M. Enkin, and M. J.N.C.Keirse (eds.) (New York, NY: Oxford University Press, 1989).
107. Chalmers, T. C., "Problems Induced by Meta-Analyses," *Statistics in Medicine 10(6):971-979, 1991.*
*108.* Chalmers, I., Cochrane Centre, Oxford, England, personal communication, Feb. 10, 1994.
109. Chalmers, T. C., "Screening for Breast Cancer: What Should National Health Policy Be?" (editorial), *Journal of the National Cancer institute 85(20): 1619-1624, 1993.*
110. Chalmers, I., Enkin, M., and Keirse, M. J, N. C., "Preparing and Updating Systematic Reviews of Randomized Controlled Trials of Health Care," *Milbank Quarterly* 71(3):41 *1-437,* 1993.
111. Chalmers, T. C., Berrier, J., Sacks, H. S., et al., '* Meta-Analysis of Clinical Trials as a Scientific Discipline 11: Replicate Variability and Comparison of Studies that Agree and Disagree," *Statistics in Medicine* 6:733-744, 1987.
112. Chalmers, T. C., Hewett, P., Reitman, D., et al., "Selection and Evaluation of Empirical Research in Technology Assessment," *International Journal of Technology Assessment in Health Care* 5(4):52 1-536, *1989.*
113. Chalmers, T. C., Levin, H., Sacks, H. S., et al., "Meta-Analysis of Clinical Trials as a Scientific Discipline. I: Control of Bias and Comparison with Large Co-Operative Trials, " *Statistics in Medicine* 6:3 *15-325, 1987.*
114. Chapin, A., President, Maine Health Information Center, Augusta, ME, personal communication, Jan. 6, 1994.
115, Chassin, M. R., ● 'Explaining Geographic Variations: The Enthusiasm Hypothesis," *Medical Care 31(5)* (suppl.):YS37-YS44, *1993.*
116. Chassin, M. R., and MeCue, S. M., "A Randomized Trial of Medical Quality Assurance," *Journal of the American Medical Association 256: 1012-1016, 1986.*
117. Chassin, M. R., Brook, R. H., Park, R. E., et al., "Variations in the Use of Medical and Surgical Services by the Medicare Population," *New England Journal of Medicine* 314:285-290, *1986.*
118. Chassin, M, R., Kosecoff, J., Park, R. E., et al., "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services?" *Journal of the American of Medical Association 258(1* 8):2533-2537, *1987.*
*119.* Cherkin, D. C., Deyo, R. A., Wheeler, K., et al., "Physician Variation in Diagnostic Testing for Low Back Pain," *Arthritis and Rheumatism 37(l): 15-22, 1994.*
*120.* Chokshi, A. B., Friedman, H. S., Malach, M., et al., "Impact of Peer Review in Reduction of Permanent Pacemaker Implantations," *Journal of the American Medical Association 246:754-757, 1991.*
*121.* Ciccone, K, R., and Chesnut, T. J., Hospital Association of New York State, *Medical Technology Assessment: A Model for Decision Making,* First Edition (Rockville, MD: Bader & Associates, 1992).
122. Clancy, C., Director, Division of Primary Care, Agency for Health Care Policy and Research, Rockville, MD, personal commun icat ion, March 1994.
123. Clarfield, A. M., "Standards, Guidelines and Clinical Policies," *Canadian Medical Association Journal* 147(4):402, *1992.*
*124.* Classen, D. C., Evans, R. S., Pestotnik, S. L., et al., "The Timing of Prophylactic Administration of Antibiotics and the Risk of Surgical-Wound Infection," *New England Journal of Medicine* 326(5):28 *1-286, 1992.*
125. Classen, D. C., Assistant Professor of Internal Medicine, School of Medicine, University of Utah, Salt Lake City, UT, personal communication, May 25, 1994.

126. Clinton, J. J., Administrator, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, "Agency for Health Care Policy and Research," presentation at '. Health Services Research: Implications for Policy, Management and Clinical Practice," Association for Health Services Research & Foundation for Health Services Research, Washington, DC, June 27, 1993.

127. Clinton, J. J., .'Financing Medical Effectiveness Research: Role of the Agency of Health Care Policy and Research," *Doing More Good than Harm: The Evaluation of Health Care Interventions*, K.S. Warren and F. Mosteller (eds. ) (New York, NY: New York Academy of the Sciences, 1993).

128. Coates, V., Vice President for Information Systems and Technology Assessment, Lemer, J. C., Vice President for Strategic Planning, and Opher, D., Research Assistant, Emergency Care Research Institute, Plymouth Meeting, PA, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 22, 1994.

129. Cochi, S.L., Broome, C. V., and Hightower, A. W., "Immunization of U.S. Children with *Hemophilus influenzae* Type b Polysaccharide Vaccine: A Cost-Effectiveness Model of Strategy Assessment," *Journal of the American Medical Association* 253(4):52 *1-529, 1985.*

130. Cochrane, A., *Effectiveness and Efficiency: Random Reflections on Health Services* (London, England: The Nuffield Provincial Hospitals Trust, 1972).

131. Cochrane Collaboration, "The Cochrane Collaboration: Preparing, Maintaining and Disseminating Systematic Rev iews of the Effects of Health Care," Oxford, England, brochure, undated.

132. Cockett, A., President, American Urological Association, Baltimore, MD, personal communication, June 1993.

133. Cohen, D.I. and Neuhauser, D., "The Metro Firm Trials: An Innovative Approach to Ongoing Randomized Clinical Trial s," *Assessing Medical Technology*, Institute of Medicine (cd, ) (Washington, DC: National Academy Press, 1985).

134. Cohen, S. J., Stookey, G. K., Katz, B. P., et al., "Encouraging Primary Care Physicians to Help Smokers Quit, A Randomized Controlled Trial," *Annals of Internal Medicine 110:648-52, 1989.*

135. Colburn, D., "Bone Marrow Transplants: A Tough Choice," *Washington Post*, Health Section, pp. 10, 11, 13, Apr. 19, 1994.

136. Colditz, G. A.. Miller, J. N., Mosteller, F., "How Study Design Affects Outcomes in Comparisons of Therapy, I: Medical," *Statistics in Medicine* 8(4):441 *-454,* 1989.

137. Collins, R., and Julian, D., "British Heart Foundation Surveys (1 987 and 1989) of United Kingdom Treatment Policies for Acute Myocardial Infarction," *British Heart Journal* 66:250-255, *1991.*

138. "Commentary from the Ophthalmology Community: Definition of Guidelines and Practice Patterns," *Abstracts of Clinical Cure Guidelines 5(7):* 16-18, 1993.

139. Commonwealth of Australia, *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee: Includin, Submissions Involving Economic Analyses* (Canberra, Australia: Department of Health, Housing, and Community Services, 1990, revised August 1992).

140. Concato, J., Horwitz, R. I., Feinstein, A. R., et al., *'Problems of Comorbidity in Mortality after Prostatectorny," *Journal of the American Medical Association 267: 1077-1082,* 1992.

141. "Consensus Conference on Corticosteroids' Perinatal Effect Planned," *Public Health Reports 109(* 1): *150,* 1994.

142. Concorde Coordinating Committee, "Concorde: MRC/ANRS Randomised Double-Blind Controlled Trial of Immediate and Deferred Zidovudine in Symptom-Free HIV Infection," *Lancet* 343:871 *-881, 1994.*

143. Cook, D. J., Guyatt, G. H., Ryan, G., et al., "Should Unpublished Data be Included in Meta-Analyses?" *Journal of the American Medical Association 269(2* 1):2749-2753, *1993.*

144. Cook, D. J., Guyatt, G. H., Laupacis, A., et al., "Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents," *Chest* 102(suppl.):305S-311 S, October 1992.

145. Crowder, R.E., Bills, E. S., and Bradbent, J. S., "The Management of Threatened Abortion: A Study of 100 Cases," *American Journal of Obstetrics and Gynecology* 60:896-899, *1950.*

146. Cumock, D. A., "Identifying Hearing Impairment in Infants and Young Children" (editorial), *British Medical Journal 307(6914): 1225-1226, 1993.*

*147.* Daddario, E. Q., quoted in Hettman, F., *Society and the Assessment of Technology,* (Paris, France: Organisation for Economic Co-operation and Development, 1973).

148. Dalence, C. R., "Human Papillomavirus DNA Testing in the Management of Cervical Neoplasia," *Journal of the American Medical Association 270(24):2975-2981, 1993.*

*149.* Daniels, M., and Schroeder, S. A., "Variation Among Physicians in Use of Laboratory Tests II: A Relation to Clinical Productivity and Outcomes of Care," *Medical Care* 15(6):482-487, *1977.*

*150.* Davey Smith, G., and Pekkanen, J., "Should There Be a Moratorium on the Use of Cholesterol Lowering Drugs?" *British Medical Journal* 304:43 1-434, 1992.

151. Davey Smith, G., Song, F., and Sheldon, T. A., "Cholesterol Lowering and Mortality: The Importance of Considering Initial Level of Risk," *British Medical Journal 306: 1367-1373, 1993.*

*152.* Davidson, G., "Does Inappropriate Use Explain Small-Area Variations in the Use of Health Care Services? A Critique," *Health Services Research* 28(4):389-400, 1993.

*153.* Davis, D. A., Thomson, M. A., Oxman, A. D., et al., "Evidence for the Effectiveness of CME: A Review of 50 Randomized Controlled Trials," *Journal of the American Medical Association 268(9):* 1111-1117, 1992.

154. Davis, M. E., and Fugo, N. W., "Steroids in the Treatment of Early Pregnancy Complications," *Journal of the American Medical Association* 142:778-785, *1950.*

*155.* Dawson, N. V., "Organizing the Metro Firm System for Research," *Medical Care 29(7):* JS19-25, *1991.*

*156.* Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. G., *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes* (Glenview, IL: Scott, Foresman and Company, 1975).

157. Deming, W. E., *Common Causes and Special Causes of improvement: Stable System, Out of Crisis* (Cambridge, MA: MIT Center for Advanced Engineering Study, 1986).

158. Demlo, L. K., and Campbell, P. M., "Improving Hospital Discharge Data: Lessons from the National Hospital Discharge Survey," *Medical Care* 19: 1030-1040, 1981.

159. Detsky, A. S., Naylor, C. D., O'Rourke, K., et al., "Incorporating Variations in the Quality of Individual Randomized Trials into Meta-Analysis," *Journal of Clinical Epidemiology* 45(3):255-265, *1992.*

*160.* Detsky, A. S., "'Guidelines for Economic Analysis of Pharmaceutical Products: A Draft Document for Ontario and Canada," *Pharmacoeconomics 3(5):354-361,* 1993.

161. Deyo, R. A., "The Quality of Life, Research and Care" (editorial), *Annals of Internal Medicine* 114:695-697, 1991.

162. Diabetes Control and Complications Trial Research Group, "The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus," New *England Journal of Medicine 329(14):977-986, 1993.*

*163.* Diabetic Retinopathy Study Group, ● 'Preliminary Report on the Effects of Photocoagulation Therapy," *American Journal of Ophthalmology 81* :383-396, *1976.*

*164.* Diabetic Retinopathy Study Group, "Photocoagulation Treatment of Proliferative Diabetic Retinopathy: Clinical Application of Diabetic Retinopathy Study (DRS) Findings, DRS Report No. 8," *Ophthalmology* 88:583-90, *1981.*

*165.* Diamond, G., and Denton, T., ● 'Alternative Perspectives on the Biased Foundations of Medical Technology Assessment," *Annals of Internal Medicine* 118(6):455-464, 1993.

166. Dickersin, K., "Report from the Panel on the Case for Registers of Clinical Trials at the Eighth Annual Meeting of the Society for Clinical Trials," *Controlled Clinical Trials* 9:76-8 1, 1988.

167. Dickersin, K., "Keeping Posted: Why Register Clinical Trials'? — Revisited," *Controlled Clinical Trials* 13: 170-177. 1992.

168. Dickersin, K., discussion comments at workshop on "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," Office of Medical Applications of Research, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, Dec. 6, 1993.

169. Dickersin, K., Professor, Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, MD, personal communication, March 1994.

170. Dickersin, K., and Berlin, J., "Me(a-Analysis: State-of-the-Science," *Epidemiologic Reviews* 14: 154-176, 1992.

171. Dickersin, K., Higgins, K., Meinert, C. L., "Identification of Meta-Analyses. The Need for Standard Terminology," *Controlled Clinical Trials* 11( 1):52-66, 1990.

172. Dickersin, K.. and Min, Y.]., "NIH Trials and Publication Bias," *Online Journal of Current Clinical Trials* [serial online], Dec. No 50, Apr. 28, 1993.

173. Dickersin, K., Scherer, R., and Lefebvre, C., "Identification of Relevant Studies for Systematic Rev iews," *British Medical Journal,* in press.

174. DiDomenico, R., Vice President for Public Affairs, Pfizer Hospital Products Group, New York, NY, personal communication, July 8, 1993.

175. Dieckman, W. J., Davis, M. E., Rynkiewicz, L. M., et al.. "Does the Administration of Diethylstilboestrol During Pregnancy Have Therapeutic Value?" *American Journal of Obstetrics and Gynecology* 66: 1062-1075, 1953.

176. Dingell, J. D., Representative, House of Representatives, U.S. Congress, Washington, DC, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Dec. 22, 1992.

177. Dolan, J. G., Bordley, D. R., and Miller, H., "Diagnostic Strategies in the Management of Acute Upper Gastrointestinal Bleeding: Patient and Physic ian Preferences. " *Journal of General Internal Medicine* 8(10):525-529, October 1993.

178. Dragalin, D., Towers Perrin, Rosslyn, VA, personal communications, November 1993 and April 1994.

179, Drake, B. G., Callahan, C. M., Dittus, R. S., et al., "Global Rating Systems Used in Assessing Knee Replacement Outcomes," *Journal of Arthroplasty,* in press.

180. Drazin, Y. S., Director, Drug Programs Branch, Ontario Ministry of Health, Toronto, Ontario, personal communication, Feb. 10, 1993.

181. Drummond, M. F., "Survey of Cost-Effectiveness and Cost-Benefit Analyses in Industrialized Countries," *World Health Statistics Quarterly* 38(4):383-401, *1985.*

*182.* Drummond, M. F., "Basing Prescription Drug Payment on Economic Analysis: The Case of Australia," *Health Affairs* 11(4): *19'7-201, 1992.*

183, Drummond, M. F., "Methodological Principles for Economic Evaluation of Pharmaceuticals," *British Journal of Medical Economics 6B:* 1-18, 1993.

184. Drummond, M, F., Bloom, B. S., Carrin, G., et al., "Issues in the Cross-National Assessment of Health Technology," *International Journal of Technology Assessment in Health Care* 8(4):671-682, *1992.*

*185.* Drummond, M., Brandt, A., Luce, B., et al., "Standardizing Methodologies for Economic Evaluation in Health Care: Practice, Problems, and Potential," *International Journal of Technology Assessment in Health Care* 9(1):26-36, *1993.*

*186.* Drummond, M., and O'Brien, B., "Clinical Importance, Statistical Significance and the Assessment of Economic and Quality-of-Life Outcomes," *Health Economics* 2:205-212, 1993.

*187.* Drummond, M. F., Stoddart, G. L., and Torrance, G. W., *Methods for the Economic Evaluation of Health Care Programmed* (Oxford, England: Oxford University Press, 1987).

188. Dubinsky, M., and Ferguson, J. H., "Analysis of the National Institutes of Health Medicare Coverage Assessment," *International Journal of Technology Assessment* in *Health Care* 6(3):480-488, *1990.*

189. Durand-Zaleski, I., Bonnet, F., Rochant, H., et al., "Usefulness of Consensus Conferences:

The Case of Albumen," *Lancet 340: 1388-1390, 1992.*

*190.* Durlak, J. A., and Lipsey, M. W., "A Practitioner's Guide to Meta-Analysis," *American Journal of Community Psychology 19(3):291-332, 1991.*

*191.* Dutcher, G., "Examples of U.S. Registries: AIDS Trial/AIDS Drugs," presentation at workshop on "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," Office of Medical Applications of Research, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, Dec. 6, 1993.

192. Dyck, F. J., Murphy, F. A., Murphy, J. K., et al., "Effect of Surveillance on the Number of Hysterectomies in the Province of Saskatchewan," *New England Journal of Medicine* 296: 1326-1328, 1977.

193. Eagle, K. A., Mulley, A. G., Skates, S. J., et al., "Length of Stay in the Intensive Care Unit: Effects of Practice Guidelines and Feedback," *Journal of the American Medical Association* 264:992-997, *1990.*

193a. Easterbrook, P. J., Berlin, J. A., Gopalan, R., et al., "Publication Bias in Clinical Research," *Lancet* 337(8746):867-872, *1991.*

*194.* Ebert, T. H., "Initiating the Firm System at a Community Teaching Hospital," *Medical Care 29(7)*:JS65-70, *1991.*

*195.* Emergency Care Research Institute, draft report on HDC/ABMT, as cited in Colburn, D., "Bone Marrow Transplants: A Tough Choice," *Washington Post*, Health Section, pp. 10,11,13, Apr. 19, 1994.

196. Eddy, D. M., "Variations in Physician Practice: The Role of Uncertainty," *Health Affairs* 3(2):75-89, *1984.*

*197.* Eddy, D. M., "An Introduction to a Bayesian Method for Me(a-analysis: The Confidence Profile Method," *Medical Innovation at the Crossroads, Volume I: Modern Methods of Clinical Investigation,* A. Gelijns (cd.) (Washington, DC: National Academy Press, 1990).

198. Eddy, D. M., "Clinical Decision Making: From Theory to Practice—Anatomy of A Decision," *Journal of the American Medical Association* 263(3):441-443, *1990.*

*199.* Eddy, D. M., "Connecting Value and Costs: Whom Do We Ask, What Do We Ask Them?" *Journal of the American Medical Association 264(13): 1737-1739, 1990.*

*200.* Eddy, D. M., "Designing a Practice Policy—Standards, Guidelines, and Options," *Journal of the American Medical Association* 263(22):3077-3084, *1990.*

*201,* Eddy, D. M., "Guidelines for Policy Statements: The Explicit Approach," *Journal of the American Medical Association 263(16):2239-2243, 1990.*

*202.* Eddy, D.M., "Practice Policies: Guidelines for Methods," *Journal of the American Medical Association 263(1 3): 1839-1841, 1990.*

*203.* Eddy, D. M., "Practice Policies: What Are They?" *Journal of the American Medical Association* 263(6):877-880, *1990.*

*204.* Eddy, D. M., "Practice Policies: Where Do They Come From?" *Journal of the American Medical Association 263(9): 1265, 1269, 1272, 1275, 1990.*

*205.* Eddy, D. M., "Resolving Conflicts in Practice Policies," *Journal of the American Medical Association 264(3):389-391, 1990.*

*206.* Eddy, D. M., ● 'The Individual vs. Society: Resolving the Conflict," *Journal of the American Medical Association 265(1* 8):2399-2401, *2405-2406, 1991.*

*207.* Eddy, D. M., *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach* (Philadelphia, PA: American College of Physicians, 1992).

208. Eddy, D. M., "Medicine, Money, and Mathematics," *Bulletin of the American College of Surgeons* 77(6):36-49, *1992.*

*209.* Eddy, D. M., Hasselblad, V., Shachter, R., "An Introduction to a Bayesian Method for Meta-Analysis: The Confidence Profile Method," *Medical Decision Making* 10(1): 15-23, 1990.

210. Eddy, D. M., Hasselblad, V., Shachter, R., *Meta-Analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence* (Boston, MA: Academic Press, Inc., 1992).

211. Eichhorn, J. H., "Prevention of Intraoperative Anesthesia Accidents and Related Severe Injury Through Safety Monitoring," *Anesthesiology 70:567-571, 1989.*

*212.* Eichhorn, J. H., Cooper, J. B., Cullen, D. J., et al., "Standards for Patient Monitoring During Anesthesia at Harvard Medical School," *Jour-*

*nal of the American Medical Association* 256(8):1() 17-1020, 1986.

213. Einarson, T. R., Leeder, J. S., Koren, G., "A Method for Meta-Analysis of Epidemiological Studies," *Drug Intelligence and Clinical Pharmacy 22(*10):813-824, *1988.*

*214.* Eisenberg, J. M., *Doctors Decisions and the Cost of Medical Care: The Reasons for Doctors Practice Patterns and Ways to Change Them* (Ann Arbor, MI: Health Administration Press, 1986).

215. Eisenberg, J. M., "Why a Journal of Pharmacoeconomics?" *PharmacoEconomics* 1 ( 1 ):2-4, 1992.

216. Eisenberg, J. M., and Foster, N. E., "Using Clinical Guidelines: Impact on Public Policy, Clinical Policy, and Individual Care Decisions," *Journal on Quality Improvement 19(9):384-391,* September 1993.

217. Eisenberg, J. M., Schulman, K. A., Glick, H., et al., "Pharmacoeconomic: Economic Evaluation of Pharmaceuticals," *Pharmacoeconomics*, 2nd cd., B .L. Strom (cd. ) (New York, NY: John Wiley & Sons, in press).

218. Eisenberg, J. M., Schwart, J. S., McCaslin, F. C., et al., "Substituting Diagnostic Services. New Tests only Partly Replace Older Ones," *Journal of the American Medical Association* 262(9): *196-200, 1989.*

*219.* Elixhauser, A., Luce, B. R., Taylor, W. R., et al., "Health Care CBA/CEA: An Update on the Growth and Composition of the Literature," *Medical Care* 31 (7)(suppl.):JS1-JS 11, 1993.

220. Ellenberg, S. S., "Meta-Analysis: The Quantitative Approach to Research Review," *Seminars in Oncology* 15(5):472-481, 1988.

221. Ellenberg, S. S., "DO Large, Simple Trials Have a Place in the Evaluation of AIDS Therapies?" *Oncology* 6:55-59, 1992.

222. Ellenberg, S. S., Cooper, E., Eigo, J., et al., "Studying Treatments for AIDS: New Challenges for Clinical Trials—A Panel Discussion at the 1990 Annual Meeting of the Society for Clinical Trials," *Controlled Clinical Trials* 13:272-292, 1992.

223. Ellenberg, S. S., and Foulkes, M. A., "The Utility of Large Simple Trials in the Evaluation of AIDS Treatment Strategies," *Statistics in Medicine* 13:405-413, 1994.

224. Ellison, L., Database Manager, Prevention Guidelines Database, Public Health Programs Office, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Serv ices, Atlanta, GA, personal communication, Jan. 27, 1994.

225. Epstein, D. L., Krug, J. H., Hertzmark, E.. et al., "A Long-Term Clinical Trial of Timolol Therapy Versus No Treatment in the Management of Glaucoma Suspects," *Ophthalmo logy* 97(4): 397, April 1990.

226, Epstein, A. M., Begg, C. B., and McNeil, B. J., "The Effects of Physicians' Training and Personality on Test Ordering for Ambulatory Patients," *American Journal of Public Health 74(* 11): 1271-1273, 1984.

227. Epstein, A. M., Hall, J. A., Tognetti, J., et al., "Using Proxies to Evaluate Quality of Life: Can they Provide Information about Patients' Heath Status and Satisfaction with Medical Care?" *Medical Care* 27(3 ):s91-98, 1989.

228. EuroQol Group, "EuroQol—A New Facility for the Measurement of Health-Related Quality of Life," *Health Policy 16:* 199-208, 1990.

229. Evidence-Based Working Group, "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medic ine," *Journal of the American Medical Association* 268:2420-2425, *1992.*

*230.* Farlow, M., Gracon, S.1.. Hershey, L. A.. et al., "A Controlled Trial of Tacrine in Alzheimer's Disease," *Journal of the American Medical Association 268(* 18):2523-2529, 1992.

231, Feldstein, P., Wickizer, T., and Wheeler, J., "Private Cost Containment: The Effects of Utilization Review Programs on Health Care Use and Expenditures," *New England Journal of Medicine 3 18:13 IO-13 14, 1988.*

*232,* Felsenthal, E., "Once Hurt, People Can Be Compensated Twice Under Double-Recovery Rules," *Wall Street Journal*, p. B 1. Mar. 4, 1994.

233. Ferguson, J. H., "Effect of Stilboestrol on Pregnancy Compared to the Effect of Placebo, " *American Journal of Obstetrics and Gyneco-g} 65:592-601, 1953.*

*234.* Ferguson, J. H.. presentation at "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," conference sponsored by

National Institutes of Health, Public Health Service, U.S. Department of Health, Rockville, MD, Dec. 6-7, 1993.

235. Ferguson, J. H., "Background and Goals of the Workshop," presentation at workshop on "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," Office of Medical Applications of Research, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, Dec. 6, 1993,

236. Ferguson, J. H., "NIH Consensus Conferences: Dissemination and Impact," *Annuls of the New York Academy of Sciences*, 703: 180-98, Dec. 31, 1993,

237. Ferguson, J. H., Director, Office of Medical Applications of Research, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, personal communication, Jan. 26, 1994, and Feb. 1, 1994.

238. Fine, M. J., Singer, D. E., Phelps, A. L., et al., "Differences in Length of Hospital Stay In Patients with Community-Acquired Pneumonia: A Prospective Four-Hospital Study," *Medical Cure* 31(4):371-380, 1993.

239. Fine, M. J., Singer, D. E., Hanusa, B. H., et al., "bValidation of a Pneumonia Prognostic Index Using the MedisGroups Comparative Hospital Database," *The American Journal of Medicine* 94: 153-159, 1993.

240. Fineberg, H. V., "Effects of Clinical Evaluation on the Diffusion of Medical Technology," *Assessing Medical Technologies*, Institute of Medicine (cd.) (Washington, DC, National Academy Press, 1985).

241. Fineberg, H. V., "Technology Assessment. Motivation, Capability, and Future Directions," *Medical Cure 23(5):663-671, 1985.*

*242.* Fineberg, H. V., and Pearlman, L. A., "Low-Cost Medical Practices," *Annual Review of Public Health,* vol. 3, L. Breslow (cd.) (Palo Alto, CA: Annual Reviews, Inc., 1982),

243. Fineberg, H. V., Gabel, R., and Sosman, M., "The Acquisition and Application of New Medical Knowledge by Anesthesiologists: Three Recent Examples," *Anesthesiology})* 48:430-436, *1978.*

*244.* Fineberg, H. V., and Hiatt, H. H., "Evaluation of Medical Practices: the Case for Technology Assessment," *New England Journal of Medicine 301(20): 1086-1091, 1979.*

*245.* Fink, A., Brook, R. H., Kosecoff, J., et al., "Sufficiency of Clinical Literature on the Appropriate Uses of Six Medical and Surgical Procedures," *Western Journal of Medicine* 147(5):609-614, *1987.*

*246.* Fink, A., Kosecoff, J., Chassin, M., et al., "Consensus Methods: Characteristics and Guidelines for Use," *American Journal of Public Health* 74(9):979-983, *1984.*

*247.* Fisher, E. S., Malenka, D. J., Wennberg, J. E., et al., "Technology Assessment Using Insurance Claims. Example of Prostatectomy," *International Journal of Technology Assessment in Health Care 6(2): 194-202, 1990.*

*248.* Fisher, E. S., and Wennberg, J. E., "Administrative Data in Effectiveness Studies: The Prostatectomy Assessment," *Proceedings; Effectiveness and Outcomes in Health Care,* K.N. Lohr and K.A. Heithoff (eds.) (Washington, DC: National Academy Press, 1990).

249. Fisher, E. S., Whaley, F. S., Krushat, W. M., et al., "The Accuracy of Medicare's Hospital Claims Data: Progress Has Been Made but Problems Remain," *American Journal of Public Health* 82:243-248, 1992.

*250.* Flamm, B. L., *Birth After Cesarean: The Medical Facts (New* York, NY: Prentice Hall Press, 1990).

251. Fleiss, J. L., and Gross, A, J., "Meta-Analysis in Epidemiology, with Special Reference to Studies of the Association Between Exposure to Environmental Tobacco Smoke and Lung Cancer: A Critique," *Journal of Clinical Epidemiology 44(2): 127-139, 1991.*

*252.* Fleming, B., "AHCPR Guideline Approach," presentation at the workshop, *National Heart, Lung, and Blood Institute Clinical Guidelines: Planning for the Future,* Rockville, MD, Feb. 22, 1994.

253. Fleming, C., Wasson, J. H., Albertsen, P. C., et al., *'A Decision Analysis of Alternative Treatment Strategies for Clinically Localized Prostate Cancer," *Journal of the American Medical Association* 269(20):2650-2658, *1993,*

*254.* Fletcher, A., Gore, S., Jones, D., '*Quality of Life Measures in Health Care, II: Design, Analysis, and Interpretation," *British Medical Journal 305: 1145-1148, 1992.*

255. Fletcher, S. W.. Black, W., Harris, R., et al., "Report of the International Workshop on Screening for Breast Cancer," *Journal of the National Cancer Institute 85(20):* 1644-1656, 1993.

256. Flood, A. B., Scott, W. R., and Ewy, W., "Does Practice Make Perfect? Part I: The Relation Between Hospital Volumes and Outcomes for Selected Diagnostic Categories," *Medical Cure* 22(2):98- 114.1984.

257. Flood, A. B., Scott, W. R., and Ewy, W.. "Does Practice Make Perfect'? Part II: The Relation Between Volume and Outcome and Other Hospital Characteristic s," *Medical Cure 22(2) 115-125,* 1984.

258. Flynn, K., Reinhard, J., Koller, C. F., et al., "Tech hnology Management Information Exchange (TEM INEX ): Applied Technology Assessment for HMOs," *HMO Practice 6(2):* 18-24, 1992.

259. Foote, S. B., "Assessing Medical Technology Assessment: Past, Present, Future," *Milbank Quarterly 65(1 ):*59-80, *1987.*

*260.* Ford, L.. Director, Community Clinical Oncology Program, National Cancer Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication. Mar. 7, 1994.

261. Ford, L. G., Hunter, C. P., Diehr, P., et al., "Effects of Patient Management Guidelines on Physician Practice Patterns: The Community Hospital Oncology Program experience," *Journal of Clinical Oncology* 5:504-5 11,1987,

262. Fowkes, F. G.. Evans, R. C., Williams, L. A., et al., "Implementation of Guidelines for the Use of Skull Radiographs in Patients With Head Injuries," *Lancet* 2:795-796, 1984.

263. Fowler, F. J., "Using Patients' Reports to Evaluate Medical Outcomes," paper prepared under contract for the Office of Technology Assessment, U.S. Congress, Washington, DC, forthcoming 1994.

264. Fowler, F. J., Barry, M. J., Lu-Yao, G., et al., "Patient-Reported Complications and Follow-Up Treatment After Radical Prostatectomy," *Urology* 42(6):622-629, 1993.

265. Fox, R. D., Mazmanian, P. E.. Putnam, R.W. (eds. ), *Changing and Learning in the Lives of Physicians* (New York, NY: Praeger, 1989).

266. Francis, T., Koms, F. T., Voight, R. B., et al., "An Evaluation of the 1954 Poliomyelitis Vaccine Trials: Summary Report," *American Journal of Public Health* 45:1-49,1955.

*267.* Fraser, E. J., Grimes, D. A.. and Schulz, K. F., "Immunization as Therapy for Recurrent Spontaneous Abortion: A Review and Meta-Analy - sis," *Obstetrics and Gynecology 82(5):854-859, 1993.*

*268.* Froberg, D., and Kane, R., 'methodology for Measuring Health-State Preferences—Pt. I: Measurement Strategies," *Journal of Clinical Epidemiology* 42(4):345-354, *1989.*

*269.* Furberg, C. D., and Morgan, T. M., "Lessons from Overv iews of Cardiovascular Trial s," *Statistics in Medicine 6(3):295-306, 1987.*

*270.* Garber, A. M., "Can Technology Assessment Control Health Care Expenditures?" paper presented at American Economic Association annual meeting, Anaheim, CA, Jan. 5, 1993.

271. Gardner, E., "What Comes First, Data or Guidelines?" *Modern Healthcare* 22(36):28-29, September 1992.

272. Gamier, H. S., Flamant, R., and Fohanno, C., "Assessment of the Role of Randomized Controlled Trials in Establishing Treatment Polic ies," *Controlled Clinical Trials* 3:227-234, *1982.*

*273.* Gehan, E. A., "Progress in Therapy in Acute Leukemia: Randomized vs. Nonrandomized Clinical Trials," *Controlled Clinical Trials 3: 199-208, 1982.*

*274.* Gelber, R. D., and Goldhirsch, A., "Meta-Analysis in Clinical Research" (letter), *Annals of Internal Medicine 108(1): 158-159,* January 1988.

275. Gellhorn, A., "Periodic Physician Recredentialing," *Journal of the American Medical Association* 265(6):752-755, *1991.*

*276.* Gerding, R. L., Emerman, D. L., Effron, D., et al., "Outpatient Management of Partial-Thickness Bums: Biobrane versus 1% Silver Sulfadiazine," *Annuls of Emergency Medicine 19: 121-129, 124-132, 1990.*

*277. Gibson,* P., "Asthma Guidelines and Evidence-Based Medicine" (letter). *Lancet* 342:1305, 1993.

278. Gilbert. J., McPeek, B.. Mosteller, F., "Progress in Surgery and Anesthesia: Benefits and Risks of Innovative Therapy," *Costs, Risks, and*

*Benefits of Surgery,* J.P. Bunker, B.A. Barnes, and F. Mosteller (eds.) (New York, NY: Oxford University Press, 1977).

279. Gitman, L., and Koplowitz, A., "Use of Diethylstilboestrol in Complications of Pregnancy," New *York State Journal of Medicine* 50:2823-2924, *1950.*

*280.* GIVIO Investigators, "Impact of Follow-up Testing on Survival and Health-Related Quality of Life in Breast Cancer Patients," *Journal of the American Medical Association 27 1(20): 1587-1592, 1994.*

*281.* Glass, G. V., "Primary, Secondary, and Meta-Analysis of Research," *Educational Researcher* 5:3-8, *1976.*

*282.* Gleeson, S., Blue Cross and Blue Shield Association, Chicago, IL, personal communication, November 1993.

283. Gleicher, N., "Cesarean Section Rates in the United States: The Short-Term Failure of the National Consensus Development Conference in 1980," *Journal of the American Medical Association* 252:3273-3276,1984.

*284.* Global Utilization of Streptokinase and Tissue Plasmogen Activator for Occluded Coronary Arteries (GUSTO) Investigators, "An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial In farction," New *England Journal of Medicine* 329(10):673-682, *1993.*

*285.* Glover, J. A., "The Incidence of Tonsillectomy in School Children," *Proceedings of/he Royal Society of Medicine 31: 1219-1235, 1938.*

*286.* Goldberg, H. I., Mullen, M., Ries, R. K., et al., "Alcohol Counseling in a General Medicine Clinic: A Randomized Controlled Trial of Strategies to Improve Referral and Show Rates," *Medical Care 29(7)* (suppl.):JS49-JS56, July 1991.

287. Goldberg, K. C.. Hartz, A.J., Jacobsen, S. J., "Racial and Community Factors Influencing Coronary Artery Bypass Graft Surgery Rates for All 1986 Medicare Patients," *Journal of the American Medical Association 267: 1473-1477, 1992.*

*288.* Goodman, S. N., "Meta-analysis and Evidence," *Controlled Clinical Trials* 10: 188-204, 1989,

289. Gordon, B., Office of Health Technology Assessment, Agency for Health Care Pol icy and Research, Public Health Service, U.S. Department of Health and Human Services, personal communication, Rockville, MD, May 1994.

290. Gosfield, A., Alice Gosfield and Associates, Philadelphia, PA, personal communication, May 18, 1993.

291. Gottlieb, L. K., Sokol, H. N., Murrey, K. O., et al., "Algorithm-Based Clinical Quality Improvement: Clinical Guidelines and Continuous Quality Improvement," *HMO Practice* 6(1):5-12, *1993.*

*292.* Gottlieb, L. K., Associate Medical Director, Clinical Quality Management, Harvard Community Health Plan, Inc., Brookline, MA, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Mar. 7, 1994.

293. Grant, A., Valentin, L., Elbourne, D., et al., "The Effect of a Policy of Routine Formal Fetal Movement on the Counting Risk of Antepartum Late Death in Normally Formed Singleton Fetuses," *Lancet* 2:345-349, *1989.*

*294.* Grassley, C. E., Senator, U.S. Congress, Washington, DC, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, July 29, 1992.

295. Gray, B., "The Legislative Battle Over Health Services Research," *Health Affairs* 11(4):38-66, 1992.

296. Greco, P. J., and Eisenberg, J. M., "Changing Physicians' Practices," New *England Journal of Medicine 329(17):* 1271-1274, 1993.

297. Green, L. A., Miller, R. S., Reed, F. M., et al., "How Representative of Typical Practice are Practice-Based Research Networks? A Report from the Ambulatory Sentinel Practice Network Inc. (ASPN)" *Archives of Family Medicine* 2:939-949, 1993.

298. Green, L. W., Eriksen, M. P., and Schor, E. L., "Preventive Practices by Physicians: Behavioral Determinants and Potential Interventions," *American Journal of Preventive Medicine* 4(4 suppl.): 101-107, 1988.

299. Green, S. B., and Byar, D. P., "Using Observational Data From Registries to Compare Treatments: The Fallacy of Omnimetrics," *Statistics in Medicine* 3(4):361-*370, 1984.*

*300.* Greene, R. J., Director, Center for Medical Effectiveness Research, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Ser-

vices, Rockville, MD, personal communication, April 1994.

301. Greenfield, S., Director, Division of Health Services Research, New England Medical Center, Boston, MA, personal communication, February 1993.

302. Greenfield, S., Kaplan, S. H., Ware, Jr., J. E., et al., "Patients' Participation in Medical Care: Effects on Blood Sugar Control and Quality of Life in Diabetes," *Journal (!/'Gencral Internal Medicine* 3(5):448-457, *I* 988.

303. Greenland, S.. "Quantitative Methods in the Review of Epidemiologic Literature," *Epidemiologic Reviews 9: I -30,* 1987.

3(M. Greenspan, A. M., Kay, H. R., Berger, B. C., et al., "Incidence of Unwarranted Implantation of Permanent Cardiac Pacemakers in a Large Medical Population '" *New England Journal of Medicine* 31*8(3): 158-16.3,* 1988.

305. Greenwald, P., Director, Division of Cancer Prevention and Control, National Cancer Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, N! I>. personal communication, Mar. 5, 1994.

306. Greer, A.L., "The State of the Art Versus the State of the Science: The Diffusion of New Medical Technologies into Practice," *International Journal of Tec'hnology Assessment in Health Care* 4:5-26, 1988.

307. Grill i, R., and Lomas, J., 'Evaluating the Message: The Relationship Between Compliance Rate and the Subject of A Practice Guideline," *Medical Care* 32(3 ):202-2 1 3, 1994.

308. Grimshaw, J.M., and Russell, I. T., "Effect of Clinical Guidelines on Medical Practice: A Systematic Review of Rigorous Evaluations," *Lancet* 342:13 17-1322, 1993.

309. Gruppo Italiano por lo Studio della Streptochinasi nell' Infarto Miocardico (GISSI ), "Effectiveness it Intravenous Thrombolytic Treatment In Acute Myocardial Infarction," *Lanc et* 1:397-402, 1986.

31 *()*. Gruppo Italiano por lo Studio della Sopravvivenza ne I I' Infarto Miocardico, '. GISSI-2: A Factorial Randomized Trial of Alteplase versus Streptokinase and Heparin versus no Heparin among 12.49() Patients with Acute Myocardia 1 Infarction," *Lancet* 3.36:65-71, 1990.

311. Gruppo Italiano por 10 Studio della Sopravvivenza nel 1' Infarto Miocard ice, '. GISSI-3: Effects of Lisinopril and Transdermal Glyceral Trinitrate Singly and Together on 6-Week Mortality and Ventricular Function after Acute Myocardial Infarction,'" *Lancet 343:11* 15-1122. 1994.

312. "Guidelines for Doctors in the New World" (editorial), *Lancet* 339(8803): 1197-1198, 1992.

313. Gulati, S. C., and Bennett, C. L.. "Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) as Adjunct Therapy in Relapsed Hodgkin Disease," *Annals of Internal Medicine* 116(3): 177-182, 1992.

314. "GUSTO Vindicates TPA, '. *Scrip World Pharmaceutical News* 1818(19):28-30, May 7/ 11, 1993.

315. Guyatt, G. H., '.E\idencc-Based Medicine," *ACP Journal Club* 1 *12(suppl.* 2): A- 16, March/ April 1991.

316. Guyatt, G. H., Feeny, D. H.. Patrick, D. L., 'measuring Health-Related Quality of Life," *Annals of Internal Medicine 11* 8(8):622-629, 1993.

317. Hadler, S.. Director, Surveillance and Epidemiology Division, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, personal communication, Atlanta. GA, Dec. 6, 1993, and Jan. 26, 1994.

318. Hadorn, D. C., "Defining Basic Health Benefits Using Clinical Guidelines: A Model Proposal for Discuss ion," paper presented at "Designing a Fair and Reasonable Basic Benefit Pkm Using Clinical Guidelines: A California Proposal," conference sponsored by the California Public Employees Retirement System's Health Benefits Advisory Council. Sacramento, CA, Apr. 23-26, 1991.

319. Hadorn, D. C., "Necessary-Care Guidelines: Using an Explicit Standard of Proof to Define Health Care Need\," paper presented at "Designing a Fair and Reasonable Basic Benefit Plan Using Clinical Guidelines: A California Proposal," conference sponsored by the California Public Employees Retirement System 's Health Benefits Ad\ isory Council, Sacramento, CA, Apr. 23-26, 1991.

*320.* Hadorn, D. C., and Brook, R. H., "The Health Care Resource Allocation Debate: Defining Our Terms," paper presented at "Designing a Fair and Reasonable Basic Benefit Plan Using Clinical Guidelines: A California Proposal," conference sponsored by the California Public Employees Retirement System's Health Benefits Advisory Council, Sacramento, CA, Apr. 23-26, 1991.

321. Hadorn, D. C., McCormick, K., and Diokno, A., "'An Annotated Algorithm Approach to Clinical Guideline Development," *Journal of the American Medical Association* 267(24):3311-3314, 1992.

322. Hall, M. A., and Anderson, G. F., *'Health Insurers' Assessment of Medical Necessity." *University of Pennsylvania Law Review 140: 1637-1712,* 1992. -

323. Halperin, W., and Neuhauser, D., "MEU: A Way of Measuring Management Efficient Utilization of Hospital Services," *Heal/h Care Management Review* 1(2):63-70, *1976.*

*324.* Hannah, W. J., Baskett, T. F., Chance, G. W., et al., "Indications for Cesarean Section: Final Statement of the Panel of the National Consensus Conference on Aspects of Cesarean Birth," *Canadian Medical Association Journal* 134: 1348-1352, 1986.

325. Harrell, J. A., Deputy Director, Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Public Health Service, U.S. Department of Health and Human Services, Washington, DC, personal communication, Dec. 10, 1993.

326. Harris, J. S., '*Why Doctors Do What They Do," *Journal of the American Medical Association* 263(5):683-687, 1990.

*327.* Harter, J., Medical Officer, Food and Drug Administration, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, April 1993,

328. Hasley, P. B., Brancati, F. L., Rogers, J., et al., "Measuring Functional Change in Community-Acquired Pneumonia: A Preliminary Study Using the Sickness Impact Profile," *Medical Care* 31(7):649-657, 1993.

*329.* Hatch, L., Health Insurance Association of America, Washington, DC, personal communication, November 1992.

330. Haynes, R. B., Faculty of Health Sciences, Clinical Epidemiology and Medicine, McMaster University, Hamilton, Ontario, personal communication, March 1993.

331. Haynes, R. B., Davis, D. A., McKibbon, A., et al., "A Critical Appraisal of the Efficacy of Continuing Medical Education," *Journal of the American Medical Association* 251:61-64, *1984.*

*332.* Hayward, R. S. A., Steinberg, E. P., Ford, D, E., et al., "Preventive Care Guidelines: 1991," *Annals of Internal Medicine,* 11 4(9):758-783, May 1991.

333. Headrick, L. A., Speroff, T., Pelecanos, H. I., et al., "Efforts to Improve Compliance With the National Cholesterol Education Program Guidelines: Results of a Randomized Controlled Trial," *Archives of Internal Medicine* 152:2490-2496, December 1992.

334. Heinen, L., and Weatherman, S., "Quality Evaluation: A New State of the Art," *Group Practice Journal 41(* 1 ):38-43, January/February 1992.

335. Hemminki, E., "Problems of Clinical Trials as Evidence of Therapeutic Effectiveness," *Social Science and Medicine 16:711 -712, 1982.*

*336.* Henry, D. A., Wilson, A., "Meta-Analysis Part 1: An Assessment of Its Aims, Validity, and Reliability," *Medical Journal of* Australia 156(1):31-38, 1992.

337. Herbst, A. L., "DES-Associated Clear Cell Adenocarcinoma of the Vagina and Cervix," *Obstetrical and Gynecological Survey* 34:844, *1979.*

*338.* Herbst, A. L., and Scully, R. E., "Adenocarcinoma of the Vagina in Adolescence: A Report of 7 Cases Including 6 Clear Cell Carcinomas (So Called Mesonephromas)," *Cancer* 25:745-757, *1970.*

*339.* Herbst, A. L., Hubby, M. M., Blough, R. R., et al., "A Comparison of Pregnancy Experience in DES-Exposed and DES-Unexposed Daughters," *Journal of Reproductive Medicine 24:62-69, 1980.*

*340.* Herbst, A. L., Ulfeder, H., and Poskanzer, D. C., "Adenocarcinoma of the Vagina: Association of Maternal Stilboestrol Therapy With Tumor Appearance in Young Women," *New England Journal of Medicine* 284:878-881, *1971.*

*341.* Hershey, C. O., Goldberg. H. I., and Cohen, D. I., "The Effect of Computerized Feedback Coupled with a Newsletter Upon Outpatient Prescribing Charges: A Randomized Controlled Trial," *Medical Care* 26:88-93, 1988.

342. Hershey, C. O., Porter, D. K., Breslau, D.. et al., "Influence of Simple Computerized Feedback on Prescription Charges in an Ambulatory Clinic,"' *Medical Care 24:472-481,* 1986.

343. Hillborne, L. H., Leape, L. L., Bernstein, S. J., et al., "The Appropriateness of Use of Percutaneous Transluminal Coronary Angioplasty in New York State." *Journal of the American Medical Association* 269(6):761-*765, 1993.*

*344.* Hillman, A. L., Pauly, M. V., and Kerstein, J. J., "How Do Financial incentives Affect Physicians' Clinical Decisions and the Financial Performance of Health Maintenance Organizations'?" *New England Journal of Medicine* 321:86-92, 1989.

345. Hillman, B. J., Joseph, C. A., Mabry, M. R., et al., "Frequency and Costs of Diagnostic Imaging in Office Practice: A Comparison of Self-Referring and Radiologist Referring Physicians," *New England Journal of Medicine 323(23): 1604-1608, 1990.*

*346.* Hine, L. K., Laird, N.. Hewitt, P., et al., "Meta-Analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction," *Archives of Internal Medicine 149:2694-2698, 1989.*

*347.* Hinman, A. R., and Koplan, J. P., "Pertussis and Pertussis Vaccine: Reanalysis of Benefits, Risks, and Costs." *Journal of the American Medical Association 251* (23):3 109-3113, 1984.

348. Hisashige, A., "Technology Assessment of Periodic Health Examinations in School Children in Japan," *International Journal of Technology Assessment in Health Care* 8(2):325-334, *1992.*

*349.* Hlatky, M. A., Califf, R. M., Harrell, F. E., et al., "Comparison of Predictions Based on Observational Data with the Results of Randomized Controlled Clinical Trials of Coronary Artery Bypass Surgery," *Journal of the American College of Cardiology* 11 (2):237-245, *1988.*

*350.* Hoffman, R. M., Kent, D. L., and Deyo, R. A., "Diagnostic Accuracy and Clinical Utility of Thermography for Lumbar Radiculopathy," *Spine* 16(6):623-628, 1991.

351. Holohan, T., Director, Office of Health Technology Assessment, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, Jan. 25, 1993

352. Holroyd, K. A., and Penzien, D.B., "Meta-Analysis Minus the Analysis: A Prescription for Confusion," *Pain 39(3)359-363.* 1989.

*353.* Holtgrave, D. R., Valdiserri, R. O., Garber, A. H., et al., "Human Immunodeficiency Virus Counseling, Testing, Referral and Partner Notification Services: A Cost-Benefit Analysis." *Archives of Internal Medicine 153( 10): 1225-1230, 1993.*

*354.* Horbar, J. D., Database Project Director, Neonatal Research and Technology Assessment, Inc., Vermont-Oxford Trials Facilitation Service, Burlington, VT, letter to Office of Technology Assessment, U.S. Congress, Washington, DC, June 20, 1992.

355. Horwitz, R. I., "Complexity and Contradiction in Clinical Trial Research," *American Journal of Medicine* 82:498-5 10, 1987.

356. Horwitz, R. I., and Feinstein, A. R., "Improved Observational Method for Studying Therapeutic Efficacy: Suggestive Evidence that Lidocaine Prophylaxis Prevents Death in Acute Myocardial Infarction," *Journal of the American Medical Association* 246:2455-2459, *1981.*

*357.* Horwitz, R. I., Viscolli, C. M., and Clemens, J. D., et al., "Developing Improved Observational Methods for Evaluating Therapeutic Effectiveness," *American Journal of Medicine* 89:630-638, *1990.*

*358.* Hsia, D. C., Krushat, W. M., Fagan, A. B., et al., "Accuracy of Diagnostic Coding for Medicare Patients Under the Prospective Payment System," *New England Journal of Medicine* 318:352-355, 1988.

359. Hubbard, S. M., Director, International Cancer Information Center, and Kramer, B. S., Associate Director, Early Detection and Community Oncology Program, Division of Cancer Prevention and Control, National Cancer Insti-

tute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, letter to the Office of Technology Assessment, U.S. Congress, Washington. DC, Apr. 21. 1994.

360. Hunt. S. M., McEwen, J., and McKenna, S. P., *Measuring Health Status (Dover,* NH: Croom Helm, 1986).

361. Idhe, D., Deputy Director, National Cancer Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, Apr. 9,1993

362. Iezzoni, L. I., Associate Professor, Harvard University, School of Medicine, Boston, MA, personal communication, Feb. 1, 1994.

363, Iezzoni, L. I., Foley, S. M., Daly, J., et al., "Cormorbid it ies, Complications, and Coding Bias: Does the Number of Diagnosis Codes Matter in Predicting In-Hospital Mortality?" *Journal of the American Medical Association 267(* 16): *2197-2203, 1992.*

*363a.* Iglehart, J. K., "Another Chance for Technology Assessment," *New England Journal of Medicine 309(8* ):509-5 *12, 1983.*

364. "Information Has Its Price," *Modern Health-(are* 23(17):48, April 1993.

365. Institute of Medicine, *Reliability of Hospital Discharge Abstracts* (Washington, DC: National Academy Press, 1977).

366. Institute of Medicine, *Assessing Medical Technologies,* C. Goodman (cd, ) (Washington, DC: National Academy Press, 1985).

367. Institute of Medicine, *Medical Technology Assessment Directory: A Pilot Reference to Organizations, Assessments, and Information Resources, C.* Goodman (cd. ) (Washington, DC: National Academy Press, 1988).

368. Institute of Medicine, *Assessment of Diagnostic Technology in Health Care: Rationale, Methods, Problems, and Directions: Monograph of the Council on Health Cure Technology,* H. Sex, S. Stem, D. Owens, et al. (eds. ) (Washington, DC: National Academy Press, 1 989).

369. Institute of Medicine, *Controlling Costs and Changing Patient Care?: The Role of Utilization Management,* B.H. Gray and M.J. Field (eds. ) (Washington, DC: National Academy Press, 1989).

370. Institute of Medicine, *Technology Innovation: Comparing Development of Drugs, Devices, and Procedures in Medicine,* A.C. Gelijins (cd. ) (Washington, DC: National Academy Press, 1 989).

371. Institute of Medicine, *Clinical Practice Guidelines: Directions for a New Program,* M.J. Field and K. N. Lohr ( eds. ) ( Wash i ngton, DC': N ational Academy Press, I w)).

372. Institute of Medicine, *Consensus Development at NIH: Improving the Program, C.* Goodman (ed. ) (Washington, DC: National Academy Press, 1 990).

373 Institute of Medicine, *Effectiveness and Outcomes in Health Care: Proceedings* ***Of' an In-*** *vitational Conference by the in.stitute* ***O f*** *Medicine Division of Health Care Services,* K.A. Heithoff and K.N. Lohr (eds. ) (Washington. DC: National Academy Press, I! 90).

374 Institute of Medicine, *National Priorities for the Assessment of Clinical Conditions and Medical Technologies: Report of a Pilot Study,* M.E. Lara and C. Goodman (eds. ) (Washington, DC: National Academy Press, 1990).

375 Institute of Medicine, *Quality of Life and Technology Assessment: Monograph of the Council on Health Care Technology,* F. Mostel-ler and J. Falotico-Taylor (eds. ) (Washington, DC: National Academy Press, 1990).

376 Institute of Medicine, *Guidelines for Clinical Practice: From Development to Use,* M.J. Field and K .N. Lohr ( eds. ) (Washington, DC: National Academy Press, 1992).

377. Institute of' Medicine. *Setting Priorities for Health Technology Assessment* (Washington, DC: National Academy Press. 1992).

378. Institute of Medicine. Division of Health Care Services, Council on Health Care Technology, "Workshop to Improve Group Judgment for Medical Practice & Technology Assessment" (workshop materials). workshop sponsored by Institute of Medicine, Washington, DC, May 15-16.1990.

379. "Intensive Insulin Therapy for Insulin-Dependent Diabetes Mellitus" (commentary), *American College of Physicians Journal Club* 1 20(suppl. 2):30-31, March/ April 1994.

380. Internal Medicine Center to Advance Research and Ed ucat ion, *Medical Practice Guidelines*

*Workshop: issue.sf{~r- Internal Medicine (Wash*ington, DC: June 8-9, 1990).

381. International Studies of Infarct Survival (ISIS), "First Results from the Largest-Ever Heart Attack Study Show Further Survival Gains: Small, But Useful" (press release), Radcliffe Infirmary, Oxford, England., Nov. 7, 1993.

382. Inui, T. S., Yourtee, E. L., and Williamson, J. W., "Improved Outcomes in Hypertension After Physician Tutorials: A Controlled Trial," *Annals of Internal Medicine* 84(6):646-65 *1, 1976.*

*383.* Investigators of the Vermont-Oxford Trials Network Database Project, "The Vermont-Oxford Trials Network: Very Low Birth Weight Outcomes for 1990," *Pediatrics 91(3):540-545, 1993.*

*384.* ISIS-1 Collaborative Group, "Randomised Trial of Intravenous Atenolol Among 16,027 Cases of Suspected Acute Myocardial Infarction: ISIS-1," *Lancet* 2:57-66, 1986.

385. ISIS-2 Collaborative Group, "Randomised Trial of Intravenous Streptokinase, Oral Aspirin, Both, or Neither Among 17,187 Cases of Suspected Acute Myocardial Infarction: ISIS-2," *Lancet* 2:249-360, 1988.

386. *ISIS-3* Collaborative Group, "ISIS-3: A Randomised Comparison of Streptokinase vs. Tissue Plasminogen Activator vs. Anistreplase and of Aspirin Plus Heparin vs. Aspirin Alone Among 41,299 Cases of Suspected Acute Myocardial Infarction," *Lancet* 339:753-770, 1992.

387. Jachuck, S. J., Brierley, H., Jachuck, S., et al., "The Effect of H ypotensive Drugs on the Quality of Life," *Journal of the Royal College of General Practitioners 32(235): 103-105, 1982.*

*388.* Jacoby, I., "The Consensus Development Program of th, National Institutes of Health: Current Practices and Historical Perspectives," *International Journal of Technology Assessment in Health Cure* 1(2):420-432,1985.

*389.* Jacoby, I., and Pauker, S.G. 'technology Assessment in Health Care: Group Process and Decision Theory," *Israel Journal of Medical Sciences 22(3-4): 183-190, 1986.*

*390.* Jaeschke, R., and Sackett, D. L., 'bResearch Methods for Obtaining Primary Evidence," *International Journal of Technology Assessment in Health Care* 5(4):503-5 19, 1989.

391. Javitt, J. C., Tielsch, J. M., Canner, J. K., et al., "National Outcomes of Cataract Extraction: Increased Risk of Retinal Complications Associated With Nd: YAG Laser Capsulotomy," *Ophthalmology* 99: 1487-1498, 1992.

392. Javitt, J. C., Vitale, S., Canner, J. K., et al., "National Outcomes of Cataract Extraction I: Retinal Detachment after Inpatient Surgery," *Ophthalmology* 98(6):895-902, 1991.

393. Javitt, J. C., Vitale, S., Canner, J. K., et al., "National Outcomes of Cataract Extraction: Endophthalmitis Following Inpatient Surgery," *Archives of Ophthalmology 109: 1085-1089,* 1991.

*394.* Jefferson, T., and Demicheli, V., "Is Vaccination Against Hepatitis B Efficient'? A Review of World Literature," *Health Economics* 3:25-37, *1994.*

*395.* Jencks, S. F., Williams, D. K., and Kay, T. L., "Assessing Hospital-Associated Deaths from Discharge Data: The Role of Length of Stay and Cornorbid ities." *Journal of the American Medical Association* 260:2240-2246, 1988.

396. Jensen, M. P., Strom, S. E., Turner, J. A., et al., "Validity of the Sickness Impact Profile Roland Scale as a Measure of Dysfunction in Chronic Pain Patients," *Pain 50: 157-162, 1992.*

*397.* Jick, H., "The Commission on Professional and Hospital Activities Professional Activity Study: A National Resource for the Study of Rare Illnesses," *American Journal of Epidemiology* 109:625-627, 1979.

398. Jick, H., Dinan, B., and Rothman, K. J., "Oral Contraceptives and Nonfatal Myocardial Infarction," *Journal of the American Medical Association 239: 1403-1406, 1978.*

*399.* Johannesson, M., "On the Discounting of Gained Life-Years in Cost-Effectiveness Analysis," *International Journal of Technology Assessment in Health Care 8(2): 359-364,* 1992.

400. Johnson, D. H., and Brubaker, R, F.. "Glaucoma: An Overview," *Mayo Clinic Proceedings 61*:59-67, *1986.*

*401.* Johnston, M, E., Langton, K. B., Haynes, R.B. et al., 'bEffects of Computer-Based Clinical Decision Support Systems on Clinician Performance and Patient Outcome: A Critical

Appraisal of Research," *Annuls of Internal Medicine 120(2):* 135-142, 1994.

402. Juran, J. M., *Juran on Leadership for Quality: An Executive Handbook (New* York, NY: The Free Press, 1989).
403. Kahan, J. P., Bernstein, S. J., Leape, Lucian, et al., "Measuring the Necessity of Medical Procedures," *Medical Care* 32(4):357-365, *1994.*
*404.* Kahn, H., Senior Medical Director, U.S. Health Care, Philadelphia, PA, personal communication, Nov. 18, 1993.
405. Kahn, K. L., Park, R. E., Brook, R. H., et al., "The Effect of Comorbidity on Appropriateness Ratings for Two Gastrointestinal Procedures," *Journal of Clinical Epidemiology 42(2): 115-122, 1988.*
*406.* Kahn, K. L., Rogers, W., Rubenstein, L., et al., "Measuring Quality of Care Before and After Implemental ion of DRG-Based Payment," *Journal of the American Medical Association 264( 15): 1969-1973, 1990.*
*407.* Kalousdian, S., Director, Department of Technology Assessment, DATTA Program, American Medical Association, personal communication, Feb. 14, 1994 and Feb. 15, 1994.
408. Kalousdian, S., Director, Department of Technology Assessment, American Medical Association, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 121, 1994.
409. Kameen, E. M.. Assistant Attorney General, Office of the Attorney General, State of Maryland, Baltimore, MD, persona] communication and letter to the Office of Technology Assessment, U.S. Congress, Washington, DC', Jan. 6, 1994.
410. Kamlet, M. S., Carnegie Mellon University, Pittsburgh, PA, '*The Comparative Benefits Modeling Project: A Framework for Cost-Utility Analysis of Government Health Care Programs," prepared for Health Policy Studies, Office of Disease Prevention and Health Promotion, Public Health Service, U.S. Department of Health and Human Services, in cooperation with the Foundation for Health Services Research, Washington, DC, 1992.
411. Kanouse, D. E., Winkler, J. D., Kosecoff, J., et al., *Changing Medical Practice Through Technology Assessment: An Evaluation of the NIH Consensus Development Program* (Ann Arbor, MI: Association for Health Services Research and Health Administration Press, 1989).
412. Kantz, M. E., Harris, W. J., Levitsky, K., et al., "Methods for Assessing Condition-Specific and Generic Functional Status Outcomes after Total Knee Replacement," *Medical Care 30(5) (suppl.):MS240-MS252, 1992.*
413. Kaplan, R. M., Anderson, J. P., Wu, A. W., et al., "The Quality of Well-Being Scale: Applications in AIDS, Cystic Fibrosis, and Arthritis," *Medical Care 27(3) (suppl.):S27-S43, 1989.*
*414.* Kaplan, R.M. and Bush, J. W., "Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis," *Health Psychology* 1( 1 ):61-80,1982.
415. Kaplan, S. H., Greenfield, S., Dukes, K., et al., "The Effects of a Joint Physician-Patient Intervention Program on Health Outcomes and Interpersonal Care," presentation at Society of General Internal Medicine Annual Meeting, Washington, DC, Apr. 28-30, 1993.
416. Kass, M. A., Gordon, M. O., Hoff, M. R., et al., "Topical Timolol Administration Reduces the Incidence of Glaucomatous Damage in Ocular Hypertensive Individuals," *Archives of Ophthalmology 107: 1590-1598, 1989.*
*417.* Kassirer, J. P., Moskowitz, A. J., Lau, J. et al., "Decision Analysis: A Progress Report," *Annals of Internal Medicine* 106(2):275-291, 1987.
418. Katz, D., presentation at "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," conference sponsored by National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, Dec. 6-7, 1993.
419. Katz., J., *The Silent World of Doctor and Patient (New* York, NY: The Free Press, 1984).
420. Katz, J. N., Larson, M. G., Phillips, C. B., et al., "Comparative Measurement Sensitivity of Short and Longer Health Status Measurements," *Medical Care 30(* 10):9 17-925, 1992,
421. Keeler, E. B., and Cretin, S., "Uses of Cost-Benefit Analysis" (editorial), *Journal of Health Economics* 6:275 *-278,* 1987.
*422.* Keeler, E. B., and Brodie, M., "Economic Incentives in the Choice Between Vaginal Delivery and Cesarean Section," *Milbank Quarterly 71* (3):365-404, *1993.*

423. Keller, R. B., Executive Director. Maine Medical Assessment Foundation, Augusta, ME, personal communication, May 18, 1993.

424. Keller, R. B., Soule, D. N., Wennberg, J. E., et al., "Dealing With Geographic Variations in the Use of Hospitals," *Journal of Bone and Joint Surgery 72-A(9): 1286-1293.* October 1990.

425. Kelly, J. T., and Swartwout, J. E., "Development of Practice Parameters by Physician Organizations," *Quality Review Bulletin* 2:54-57, *1990.*

426. Kelly, J. V., and Hellinger, F. J., "Physician and Hospital Factors Associated With Mortality of Surgical Patients," *Medical Care* 24(9):785-800, 1986.

427. Kennedy, E. M., and Hatch, O. G., Senators, U.S. Congress, Washington, DC, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, July 18, 1992.

428. Kent, D. L., Haynor, D. R., and Larson, E. B., '*Diagnosis of Lumbar Spinal Stenosis in Adults: A Meta-Analysis of the Accuracy of CT, MR, and Myelography," *American Journal of Roentology* 158: 1135-1144, 1992.

428a. Kiefer, D. M., "Technology Assessment: A Layman's Overview," *Technology Assessment in u Dynamic Environment*, M.J. Cetron and B. Bartocha (eds.) (New York, NY: Gordon and Breach, 1973).

429. Kimball, M. C., "Joint Commission Will Launch Two Sets of Hospital Indicators This Year," *Report on Medical Guidelines and Outcomes Research* 4(2):9-10, Jan. 28, 1993.

430. Klawansky, S., Berkey, C., Shah, N., et al., "Survival from Localized Breast Cancer: Variability Across Trials and Registries," *International Journal oj' Technology Assessment in Health Cure* 9(4):539-553, *1993.*

431. Kosecoff, J., Kanouse, D. E., Rogers, W. H., et al., "Effects of the National Institutes of Health Consensus Development Program on Physician Practice," *Journal of the American Medical Association* 258:2708-2713, *1987.*

432. Kovatch, G., Committee Management Specialist, Office of the Associate Director for Science, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA, personal communication, Jan. 27, 1994.

433, Krahn, M., and Gafni, A., "Discounting in the Economic Evaluation of Health Care Interventions," *Medical Care 31* (5):403-418, 1993.

434. Kramer, M. S., and Shapiro, S. H., "Scientific Challenges in the Application of Randomized Trial s," *Journal of the American Medical Association* 252(19):2739-2745, *1984.*

435. Kravitz, R. L., Greenfield, S., Rogers, W., et al., "Differences in the Mix of Patients Among Medical Specialties and Systems of Care: Results From the Medical Outcomes Study," *Journal of the American Medical Association 267( 12): 1617-1623, 1992.*

436. L'Abbe, K. A., Detsky, A. S., and O'Rourke, K., "Meta-Analysis in Clinical Research," *Annals of internal Medicine* 107:224-233, *1987.*

437. Laird, N. M., and Mosteller, F., "Some Statistical Methods for Combining Experimental Results," *international Journal of Technology Assessment in Health Care 6(1* ):5-30, *1990.*

438. Lamas, G. A., Pfeffer, M. A., Harem, P., et al., "Do the Results of Randomized Clinical Trials of Cardiovascular Drugs Influence Medical Practice?," *New England Journal of Medicine* 327(4):241-274, *1992.*

439. Landefeld, C. S., and Anderson, P. A., "Guideline-Based Consultation to Prevent Anticoagulant-Related Bleeding: A Randomized, Controlled Trial in a Teaching Hospital," *Annals of Internal Medicine* 116:829-837, 1992.

440. Larkin, H., "Health Insurers Seek Greater Federal Role in Technology Review," *American Medical News*, Oct. 14, 1992.

441. Lasker, R. D., Shapiro, D. W., and Tucker, A. M., "Realizing the Potential of Profiling," *Conference on Profiling, No. 92-2,* Physician Review Commission (cd.) (Washington, DC: PC, 1992).

442. Lau, J., Altmann, E., Jimenez-Silva, J., et al., "Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction," *New England Journal of Medicine* 327(4):248-254, *1992.*

443. Law, M. R., Thompson, S. G., and Wald, N.J., "Assessing Possible Hazards of Reducing Serum Cholesterol," *British Medical Journal* 308:373-379, *1994.*

*444.* Leape, L. L., [b]* Practice Guidelines and Standards: An Overview," *Quality Review Bulletin* 16(2):42-49, February 1990.

445. Leape, L. L., and Brook. R. H.. *'Rand Corporation Appropriateness Rating Program: Context and Purpose," presentation at "Workshop to Improve Group Judgment for Medical Practice & Technology Assessment, an invitational workshop sponsored by Council of Health Care Technology, Division of Health Care Services, Institute of Medicine, Washington, DC, May 15-16, 1990.

446. Leape, L. L., Hilborne, L. H., Park, R. E., et al., "The Appropriateness of Use of Coronary Artery Bypass Graft Surgery in New York State," *Journal of the American Medical Association* 269(6):753-760, *1993.*

*447.* Leape, L. L., Park, R. E., Solomon, D. H., et al., "'Does Inappropriate Use Explain Small-Area Variations in the Use of Health Care Services'?" *Journal of the American Medical Association 263(5* ):669-672,. *1990.*

*448.* Leape, L. L., Park, R. E., Kahan, J. P., et al., "Group Judgments of Appropriateness: The Effect of A Panel Composition," *Quality Assurance in Health Care 4(2): 151-159,* 1992.

448a. Leary, W. E., "Agency Issues New Guidelines for Managing Heart Conditions," *New York Times* p. A20, June 29, 1994.

449. Weatherman, S., Peterson, E., Heinen, L., et al., "'Quality Screening and Management Using Claims Data In a Managed Care Setting," *Quality Review Bulletin* 17(11 ):349-359, 1991.

450. Lee, K. L., Califf, R. M., Simes, J., et al., ● 'Holding GUSTO Up to the Light," *Annals of Internal Medicine 120:876-881,* 1994.

*451.* Lee, T. H., Cook, E. F., Weisberg, M., et al., "Acute Chest Pain in the Emergency Room. Identification and Examination of Low-Risk Patients," *Archives of Internal Medicine* 145:65-69, *1985.*

*452.* Lee, T. H., Rouan, G. W., Weisberg, M. C., et al., "Sensitivity of Routine Clinical Criteria for Diagnosing Myocardial Infarction Within 24 Hours of Hospitalization," *Annals of Internal Medicine* 106: 181-186, 1987.

453. Legisletter (ACOG), '[b]Maine's Liability Demonstration Project—Relating Liability to Practice Parameters," *Legisletter* 9(4):2-5, *1990.*

*454.* Lenfant, C., Director, National Heart, Lung, and Blood Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, letter to the Office of Technology Assessment, U.S. Congress, Washington DC, April 1994.

455. Lewin, T., "Experimental Drug Is Prize in a Highly Unusual Lottery," *New York Times,* p. A-1, Jan. 7, 1994.

456. Lewis. C. E., "Variations in the Incidence of Surgery." *New England Journal of Medicine* 281:880, *1969.*

*457.* Light, R. J., and Pillemer, D. B., *Summing Up* (Cambridge, MA: Harvard University Press, 1984).

458. Light. R. J., and Smith, P. V., "Accumulating Evidence: Procedures for Resolving Contradictions Among Different Research Studies," *Harvard Education Review 41 (4):429-471, 1971.*

*459.* Lilford, R., "Formal Measurement of Clinical Uncertainty: Prelude to a Trial in Perinatal Medicine," *British Medical Journal 308:11* 1-112, *1994.*

*460.* Lindsey, P. A., and Newhouse, J. P., "The Cost and Value of Second Surgical Opinion Programs: A Critical *Review* of the Literature," *Journal of Health Politics, Policy and Law* 15(3):543-570, fall 1990.

461. Littell, C., Director, Health Care Technology Institute, Alexandria, VA, personal communications, Sept. 25, 1992 and Apr. 27, 1994

462. Llewellyn-Thomas, H. A., Sutherland, H. J., Tibshirani, R., et al., '[b]The Measurement of Patient's Values in Medicine," *Medical Decision Making* 2(4):449-462, winter 1982.

463. Llewellyn-Thomas, H. A., Sutherland, H. J., Tibshirani, R., et al., "Describing Health States: Methodologic Issues in Obtaining Values for Health States," *Medical Care* 22:543-552, 1984.

*464.* Lohr, K. N., Brook, R. H., Kamberg, C. J., et al., "Use of Medical Care in the RAND Health Insurance Experiment: Diagnosis and Service-Specific Analyses in a Randomized Trial," *Medical Care* 24(9) (suppl. ): 1-87, 19'?6.

465. Lomas, J.. "-Words Without Action'? The Production, Dissemination. and Impact of Consen-

sus Recommendation s," *Annual Review of Public Health* 12:41-65.1991.

466. Lomas, J., "Retailing Research: Increasing the Role of Evidence in Clinical Services for Childbirth," *Milbank Quarterly* 71 (3):439-475, 1993.

467. Lomas, J., and Hayrres, R. B., "A Taxonomy and Critical Review of Tested Strategies for the Application of Clinical Practice Recommendations: From 'Official' to 'Individual' Clinical Policy," *American Journal of Preventive Medicine* 4:77-94,1988.

468. Lomas, J., Anderson, G. M., Domnick-Pierre, K., et al., "Do Practice Guidelines Guide Practice? The Effect of a Consensus Statement on the Practice of Physicians," *New England Journal of Medicine 321 (*19): 1306-1311, 1989.

469. Lomas, J., Anderson, G. M., Enkin, M., et al., "The Role of Evidence in the Consensus Process: Results From a Canadian Consensus Exercise," *Journal of the American Medical Association* 259(20):3001-3005, 1988.

470. Lomas, J., Enkin, M., Anderson, G. M., et al., "Opinion Leaders vs. Audit and Feedback to Implement Practice Guidelines: Delivery After Previous Cesarean Section," *Journal of the American Medical Association 265(17):2202-2207, 1991.*

471. Long, S. H., Gibbs, J. O., Crozier, J. P., et al., "Medical Expenditures of Terminal Cancer Patients During the Last Year of Life, *Inquiry 21:3 15-327, 1984.*

472. Longnecker, M., Berlin, J., Orza, M., et al., "A Meta-Analysis of Alcohol Consumption in Relation to Risk of Breast Cancer," *Journal of the American Medical Association 260(5): 652-656, 1988.*

473. Lorrgnecker, M. P., "Meta-Analysis," paper prepared under contract for the Office of Technology Assessment, U.S. Congress, Washington, DC, forthcoming, 1994.

473a. Luce, B. R., Director, Center for Public Health Research and Evaluations, Battelle Medical Technology Assessment and Policy (MED-TAP) Research Center, Washington, DC, personal communication, June, 1993.

474. Luce, B. R., and Brown, R. E., "The Use of Technology Assessment by Hospitals, HMOS and Third Party Payers in the U.S.," *Journal of the International Society for Technology Assessment and Health Care*, in press.

475. Luce, B. R., and Simpson, K., Battelle Medica Technology Assessment and Policy (MED TAP) Research Center, Washington, DC "Methods of Cost-Effectiveness Analysis: Areas of Consensus and Debate," contract report prepared for the Pharmaceutical Manufacturers Association, Washington. DC. April 1993.

476, Luft, H. S., Bunker, J. P., and Enthoven, A. C., "Should Operations Be Regionalized? The Empirical Relation Between Surgical Volume and Mortality," *New England Journal of Medicine 301 (25): 1364-1369, 1979.*

477, Luft, H. S., Garnick, D. W., Mark, D., et al., "Evaluating Research on the Use of Volume of Services Performed in Hospitals as an Indicator of Qual it y," unpublished contractor report prepared for the Off'ice of Technology Assessment, U.S. Congress, Washington, DC, 1987.

477a, Lumsdon, K., "Beyond Technology Assessment: Balancing Needs, Strategy," *Hospitals* 66(15):20-26, 1992.

478. Lusted, L. B., "Decision Making Studies in Patient Management," *New England Journal of Medicine* 284(8):416-424, February 1971.

479, Lu-Yao, G. L., and Greenberg, E. R., "Changes in Prostate Cancer Incidence and Treatment in USA," *Lancet* 343:251-254, 1994.

480. Lu-Yao, G., McLerran, D., Wasson, J., et al., "An Assessment of Radical Prostatectomy: Time Trends, Geographic Variation, and Outcomes." *Journal of the American Medical Association* 269(20):2633-2636,1993.

481, Maass, A., Hufschmidt, M. M., Dorfman, R., et al., *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning* (Cambridge, MA: Harvard University Press, 1962).

482. Magaziner, J., Simonsick, E. M., Kashner, T. M., et al., "Patient-Proxy Response Comparability on Measures of Patient Health and Functional Status," *Journal of Clinical Epidemiology* 41(11): 1065-1074, 1988.

483. Maine Medical Assessment Foundation, *Annual Progress Report 1992* (Augusta, ME: 1993).

484. Maklan, C., Center for Medical Effectiveness Research, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, letter to Office of Technology Assessment, U.S. Congress, Washington, DC, Dec. 17, 1993.

485. Maklan, C. W., and Cummings, M. A., "Methodological Challenges and Innovations in Patient Outcomes Research," *Medical Care* (in press).

486. Mangione, C. M., Phillips, R. S., Seddon, J. M., et al., "Development of the 'Activities of Daily Vision Scale'—A Measure of Visual Functional Status," *Medical Care 30(12):11* 11-1126, 1992.

487. Mann, C., "Meta-Analysis in the Breech," Science 249:476-480, 1990.

488. Manning, P. R., Lee, P. V., Clintworth, T. A., et al., "Changing Prescribing Practices Through Individual Continuing Education," *Journal of the American Medical Association 256(2):230-232, 1986.*

*489.* Manning, W. G., Newhouse, J. P., Duan, N., et al., "Health Insurance and the Demand for Medical Care," *American Economic Review* 77(3):251-277, *1987.*

*490.* Margolis, C. Z., "Uses of Clinical Algorithms," *Journal of the American Medical Association* 249(5):627-632, *1983.*

*491.* Margolis, K. L., Lofgren, R. P., and Kern, J. E., Organizational Strategies to Improve Influenza Vaccine Delivery: A Standing Order in a General Medicine Clinic," *Archives of Internal Medicine* 148(10):2205-2207, 1988.

492. Martin, J. N., Morrison, J. C., and Wiser, W. L., "Vaginal Birth After Cesarean Section: The Demise of Routine Repeat Abdominal Delivery," *Obstetrics and Gynecology Clinics of North America* 15:719-736, *1988.*

*493.* Marwick, C., "Using High-Quality Providers to Cope With Today's Rising Health Care Costs," *Journal of the American Medical Association* 268(16):2142-2145, *1992.*

*494.* Marwick, C., "Federal Agency Focuses on Outcomes Research," *Journal of the American Medical Association 270(2): 164-165, 1994.*

*495.* Mason, J., Drummond, M., and Torrance, G., "Some Guidelines on the Use of Cost-effectiveness League Tables," *British Medical Journal* 306:570-572, *1993.*

*496.* Ma(char, D. B., "Application of Decision Analysis to Guideline Development," paper presented at Forum Methodology Conference, sponsored by Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, December 1992.

497. Materson, B. J., Reda, D. J., Cushman, W. C., et al., "Single Drug Therapy for Hypertension in Men," *New England Journal of Medicine* 328:914-921, *1993.*

*498.* Matuszewski, K. A., Director, Technology Assessment Program, University Hospitals Consortium, comments at a meeting of the Technology Assessment Task Force, National Advisory Council for Health Care Policy, Research, and Evaluation, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Resources, Sept. 22, 1993

499. Maynard, A., "The Relevance of Health Economics to Health Promotion," *World Health Organization Regional Publications European Series* 37:29-54, *1991.*

*500.* McClellan, M., and Brook, R. H., "Appropriateness of Care: A Comparison of Global and Outcome Methods to Set Standards," *Medical Care* 30(7):565-586, *1992.*

*501.* McCormick, M. C., Director, Office of the Forum for Quality and Effectiveness in Health Care, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, Jan. 19, 1993.

502. McDonald, C.J., and Overhage, J. M., "Guidelines You Can Follow and Can Trust: An Ideal and an Example," *Journal of the American Medical Association 271(1* 1):872-873, *1994.*

*503.* McDowell, I., and Newell, C., *Measuring Health: A Guide to Rating Scales and Questionnaires (New* York, NY: Oxford University Press, 1987).

504. McDowell, I., Newell, C., and Rosser, W., "Comparison of Three Methods of Recalling Patients for Influenza Vaccination," *Canadian Medical Association Journal* 135:991-997, *1986.*

*505.* McDowell, I., Newell, C., and Rosser, W., "A Randomized Trial of Computerized Reminders for Blood Pressure Screening in Primary Care," *Medical Cure* 27:297-305, *1989.*

*506.* McDowell, I. Newell, C., and Rosser, W., "Computerized Reminders to Encourage Cervical Screening in Family Practice," *Journal of Family Practice* 28:420-424, *1989.*

*507.* McGinnis, J. M., and Woolf, S. H., "Background and Objectives of the U.S. Preventive Services Task Force," *Journal of General Internal Medi*cine 5(suppl.):S11-S13, 1990.

508. McGivney, W., Director, Diagnostic and Therapeutic Technology Assessment Program, program profile in handout from "Workshop to Improve Group Judgment for Medical Practice & Technology Assessment, an invitational workshop sponsored by Council of Health Care Technology, Division of Health Care Services, Institute of Medicine, Washington, DC, May 15-16, 1990.

509. McGivney, W., Director, Technology Assessment & Policy, Aetna Health Plans, Hartford, CT, personal communication, July 21, 1993.

510. McGlynn, E. A., Kosecoff, J., and Brook, R., "Format and Conduct of Consensus Development Con ferences," *International, Journal of Technology Assessment in Health Care* 6(3):450-469, *1990.*

*511.* McGrory, D. C., Goldstein, L. B., Samsa, G. P., et al., "Predicting Complications of Carotid Endarterectomy," *Stroke 24(9):* 1285-1291, 1993.

512. McHorney, C. A., Ware, J. E., Lu, J. F. R., et al., "The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of Data Quality, Scaling Assumptions, and Reliability Across Diverse Patient Groups," *Medical Care* 32(1):40-66, 1994.

*513.* McHorney, C. A., Ware, J. E., Rogers, W., et al., "The Validity and Relative Precision of MOS Short- and Long-Form Health Status Scales and Dartmouth COOP Charts," *Medical Care 30(5) :MS253-265, 1992.*

*514.* McNeil, B. J., "Claims Data and Effectiveness: Acute Myocardial Infarction and Other Examples," *Effectiveness and Outcomes in Health Care,* K.A. Heithoff and K.N. Lohr (eds. ) (Washington, DC: National Academy Press, 1990).

515. McNeil, B, J., "Use of Claims Data to Monitor Patients over Time: Acute Myocardial Infarction as a Case Study," *Doing More Good than Harm: The Evaluation of Health Care Interventions,* K.S. Warren and F. Mosteller (eds.) (New York, NY: New York Academy of the Sciences, 1993).

516. McPhee. S. J., Bird, J. A., Fordham, P., et al., "Promoting Cancer Prevention Activities by Primary Care Physicians—Results of a Randomized, Controlled Trial," *Journal of the American Medical Association* 266:538-544, *1991.*

*517.* McPhee, S. J., Richard, R. J., and Solkowitz, S. N., "Performance of Cancer Screening in a University General Internal Medicine Practice: Comparison With the 1980 American Cancer Society Guideline s," *Journal of General Internal Medicine 1:275-281, 1986.*

*518.* Medical Alley, Technology Task Force, *Medical Alley Technology Task Force Report* (Minneapolis, MN: 1992).

519. *Medical Care,* Supplement Issue 27(3), March 1989.

520. *Medical Care,* Supplement Issue 28(1 2), December 1990.

521. "Medical Technology Assessment 'Clearinghouse' at AHCPR," *Health News Daily,* pp. 3-4, Mar. 18, 1992.

522. "Medication Adherence in Hypertension is Among AHCPR'S 12 Drug Outcome Studies," *Health News Daily,* p. *3,* Dec. *20, 1993.*

*523.* Meehan, S., Deputy Director, Health Services Research and Development Service, Veterans Health Administration, U.S. Department of Veterans Affairs, Washington, DC. personal communication, Mar. 4, 1993, and Apr. 22, 1994.

524. Merkatz, R. B., Temple, R., Sobel, S., et al., "Inclusion of Women in Clinical Trials — Policies for Population Subgroup s," *New England Journal of Medicine* 329(4):288-296, *1993.*

*525.* Merrick, N. J., Brook. R. H., Fink, A., et al., "Use of Carotid Endarterectorny in Five California Veterans Administration Medical Centers," *Journal of the American Medical Association 256(* 18):253 1-2535, 1986.

525a. Merrick, N. J., Fink, A., Brook, R. H.. et al., *Indications for Selected Medical and Surgical*

*Procedure.s-A Literature Review and Ratings of Appropriateness (Carotid Endarterectomy)* (Santa Monica, CA: RAND Corp., May 1986).

526. Merrick, N. J., Fink, A., Park, R.E. et al., "Derivation of Clinical Indications for Carotid Endarterectomy by an Expert Panel," *American Journal of Public Health 77(2): 187-190, 1987.*

*527.* Meyer, T. J., Van Kooten, D. Marsh, S., et al., "Reduction of Polypharmacy by Feedback to Clinic ians," *Journal of General Internal Medicine 6: 133-136,* 1991.

528. Mickleburgh, R., "B.C. Puts Limits on Medical Coverage," *Globe and Mail Report on Business,* p. A3, Aug. 25, 1993.

529. Miller, J. N., Colditz, G, A., and Mosteller, F., "How Study Design Affects Outcomes in Comparisons of Therapy, 11: Surgical," *Statistics in Medicine* 8(4):455-466, *1989.*

*530.* Miller-Catchpole, R., "Diagnostic and Therapeutic Technology Assessment (DATTA): Lung Transplantation," *Journal of the American Medical Association* 269(7):93 1-936, 1993.

531. Miller-Catchpole, R., "Teflon Preparations for Urinary Incontinence and Vesicouretal Reflux," *Journal of the American Medical Association* 269(23):2975-2980, *J993.*

*532.* Miller-Catchpole, R., "Diagnostic and Therapeutic Technology Assessment (DATTA): Hyperthermia as Adjuvant Treatment for Recurrent Breast Cancer and Primary Malignant Glioma," *Journal of the American Medical Association 271(* 10):797-802, 1994.

533. "Medicare Indicator System Targets Guideline Compliance," *Medical Outcomes & Guidelines Alert 2(*12):2-3, June 23, 1994.

534. Mitchell, J. M., and Sunshine, J. H.. "Consequences of Physicians' Ownership of Health Care Facilities—Joint Ventures in Radiation Therapy," *New England Journal of Medicine 327(2* 1): 1497-1506, 1992.

535. Mittman, B. S., and Siu, A. L., "Changing Provider Behavior: Applying Research on Outcomes and Effectiveness in Health Care," *Improving Health Policy and Management: Nine Critical Research issues for the 1990's,* S.M. Shorten and U.E. Reinhardt (eds.) (Ann Arbor, MI: Health Administration Press, 1992).

536. Mittman, B. S., Tonesk, X., and Jacobson, P. D., "Implementing Clinical Practice Guidelines: Social Influence Strategies and Practitioner Behavior Change," *Quality Review Bulletin* 18(12):413-422, December 1992.

537. Mohle-Boetani, J. C., Schuchat, A., Plikaytis, B. D., et al., "Comparison of Prevention Strategies for Neonatal Group B Streptoccoccal Infection: A Population-Based Economic Analysis," *Journal of the American Medical Association 270(12): 1442-1448, 1993.*

*538.* Moore, C. M., *Group Techniques for Idea Building,* Applied Social Research Methods Series, Volume 9 (Newbury Park, CA: Sage Publications, 1987).

539. Morris, L. A., "Are Standards for Economic Approaches Needed?" presentation at ● 'Pharmaceutical Economic Analyses: A Training Workshop for Government Employ ees," conference sponsored by the Center for Medical Effectiveness Research and the Center for Research Dissemination and Liaison, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, June 3, 1994.

540. Morosco, G. J., Acting Director, Office of Prevention, Education, and Control, National Heart, Lung, and Blood Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, personal communication, Feb. 15, 1994.

541. Mosteller, F., "The Contribution of Firms: A Fresh Movement in Medicine," *Medical Care 29: SJ3-4, 1991.*

*542.* Mugford, M., Banfield, P., and O'Hanlon, M., "The Effect of Computerized Feedback Coupled with a Newsletter Upon Outpatient Prescribing Charges: A Randomized Controlled Trial," *Medical Care* 26:88-93, *1988.*

*543.* Mugford, M., Banfield, P., and O'HanIon, M., "Effects of Feedback of Information on Clinical Practice: A Review," *British Medical Journal* 303:398-402, *1991.*

*544.* Muldoon, M. F., Manuck, S, B., and Matthews, K. A., "Lowering Cholesterol Concentrations and Mortality: A Quantitative Review of Primary Prevention Trials," *British Medical Journal 301*:309-3 14, 1990.

*545.* Mullan, F., and Jacoby, I., "The Town Meeting for Technology, The Maturation of Consensus

Con ferences," *Journal of the American Medical Association 254(8):* 1068, *1985.*

*546.* Muller, C.. Mandleblatt, J., and Schecter, C. B., "Costs and Effectiveness of Cervical Cancer Screening in Elderly Women" prepared for the Office of Technology Assessment, U.S. Congress, OTA-BP-H-65 ( Washington, DC: U.S. Government Printing office, February 1990).

547. Mulley, A. G., "The Role of Decision Analysis in the Translation of Research Findings into Clinical Practice," *Medical Innovation at the Crossroads, Volume 1: Modern Methods of Clinical Investigation*, A.C. Gelijns (cd. ) (Washington, DC: National Academy Press, 1990).

548. Mulley, A. G., Chief of General Internal Medicine, Massachusetts General Hospital, Boston, MA, personal communication, Oct. 4, 1993.

549. Mulley, A. G., Thibault, G. E., Hughes. R. A., et al., "The Course of Patients With Suspected Myocardial Infarction. The Identification of Low-Risk Patients for Early Transfer From Intensive Care," *New England Journal of Medifine* 302:943-948, 1980.

550. Mulrow, C. D., "The Medical Review Article: State of (he Science," *Annals of Internal Medicine 106(3):485-488,* 1987.

551. Myers, S. A.. and Gleicher, *N.,* "A Successful Program to Lower Cevwc:in-Sect ion Rates,"' *New England Journal of Medicine* 3 *19(23):* 15 I 1-1516, 1988.

552. National Association of Insurance Commissioners, Kansas City. MO, Experimental Treatments Work ing Group of the Accident and Health Insurance ( B ) Committee, minutes, June 25, 1993, Aug. 24. 1993, and Sept. 21, 1993.

553. National Health Policy Forum, George Washington Universi (y, Washington, DC, "Grading the Graders: Using 'Report Cards' to Enhance Quality of Care under Health Reform," Issue Brief No. 642, Washington, DC, Feb. 28, 1994.

554. National Science Foundation, Office of Exploratory Research and Problem Assessment, Research and Applications Directorate, *Candidates and Priorities for Assessments, Volume IV: An Approach to Priorities*, prepared by R. Ayres, A. Shapanka, and K. Humes, IRT-3 10-R (Washington, DC: (1. S. Government Printing office, March1 973).

555. Naylor, C. D., Sibbald, W. J., Sprung. C. L., et al., "Pulmonary Artery Catheterization: Can There Be an Integrated Strategy for Guideline Development and Research Promotion?" *Journal of the American Medical Association* 269(18):2407-2411, 1993.

556. The NC I Breast Cancer Screening Consortium, "Screening Mammography: A Missed Clinical Opportunity? Results of the NCI Breast Cancer Screening Consortium and National Health Interview Survey Studies," *Journal of the American Medical Association* 264:54-58,1990.

*557. Neal,* D. E., Styles, R. A., Ng, T., et al., "Relationship Between Voiding Pressures, Symptoms, and Urodynamic Findings in 253 Men Undergoing Prostatectomy," *British Journal of Urology* 60:554, 1987.

*558.* Nerenz, D. R., Repasky, D. P., Whitehouse, M. D., et al., "Ongoing Assessment of Health Status in Patients with Diabetes Mellitus Using the SF-36 and Diabetes TYPE Scale," *Medical Care* 30(5):MS 11 2-MS 124, 1992.

559. Nerenz, D. R., Zajac, B. M., M iron, L. L., et al., 'inpatient Firms in a Teaching Hospital: The Henry Ford Hospital Experience," *Medical Care* 29(7) :JS26-30, 1991.

560. Neuhauser, D., and Lewicki, A. M., "What Do We Gain from the Sixth Stool Guaiac?" *New England Journal t~'Medicine* 293(5):226-228, July 1975.

561. Neuhauser, D., "Editorial, " *Medical Cure* 24:47 1, 1986.

562. Neuhauser, D., "Ernest Avery Codman, MD, and End Results of Medical Care, " *International Journal of Technology Assessment in Health Care* 6(2):307-325, 1990.

*563.* Neuhauser, D., "Parallel Providers, Ongoing Random izat ion, and Continuous Improvement," *Medical Care* 29(7) :JS5-8, 1991.

564. Neuhauser, D., "Progress in Firms Research," *International Journal of Technology Assessment in Health Care* 8(2):32 1-324, 1992.

564a. "New Proscar/Hyrin Study in Benign Prostatic Hypertrophty (BPH)," *Scrip World Pharmaceutical News 1801: 17,* Jan. 19, 1993.

565. Newhouse, J., and the RAND Health Insurance Group, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press, 1993).

566. Nichols, S., quoted in Wagner, L., and Weissenstein, E., "Quality of Information on Plans' 'Report Cards' Getting Low Marks," *Modern Healthcare* 23(38):9, *1993.*

567. Nerd, E., "An Alternative to QALYs: the Saved Young Life Equivalent (SAVE)," *British Medical Journal* 305:875-877,1992.

568. Nerd, E., "Methods for Quality Adjustment of Life Years," *Social Science of Medicine* 34(5):599-569, *1992.*

569. Nerd, E., "The Relevance of Health State After Treatment in Prioritizing Between Different Patients," *Journal of Medical Ethics* 19:37-42, *1993.*

570. Norquist, G., Deputy Division Director and Associate Director for Services Research, Division of Epidemiology and Services Research, National Institute of Mental Health, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, fax to Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 21, 1994.

571. Nyman, J. A., Feldman, R., Shapiro. J., et al., "Changing Physician Behavior: Does Medical Review of Part B Medicare Claims Make a Difference?," *Inquiry 27(2):* 127-137, 1990.

572. O'Brien, B. J., presentation at fifth meeting of the Cost-Effectiveness Panel on Clinical Preventive Services, Office of Disease Prevention and Health Promotion, Office of the Assistant Director for Health, Public Health Service, U.S. Department of Health and Human Services, Washington DC, Mar. 8-9, 1994.

573. O'Brien, B.J., Drummond, M. F., Labelle, R.J., et al., "In Search of Power and Significance: Issues in the Design and Analysis of Stochastic Cost-Effectiveness Studies in Health Care," *Medical Care* 32(2): 150-163, 1994.

574. O'Carrel, P. W., and Potter, L. B., "Programs for the Prevention of Suicide Among Adolescents and Young Adults," *Morbidity and Mortality Weekly Reports 43( RR-6):* 1-7, April 1994.

575. Oddone, E. Z., Homer, R. D., Monger, M. E., et al., "Racial Variations in the Rates of Carotid Angiography and Endarterectomy in Patients with Stroke and Transient Ischemic Attack," *Archives of Internal Medicine* 153:2781 -*2786, 1993.*

576. Oddone, E. Z., Samsa, G., and Matchar, D. B., "Global Judgments versus Decision-Model-Facilitated Judgments: Are Experts Internally Consistent?" *Medical Decision Making 14(1): 19-26, 1994.*

577. Odone, A., and Odone, M., ● 'No One Said Lorenzo's Oil Could Prevent Rare Nerve Disease" (letter), *New York Times,* p. 20, Mar. 2, 1993.

578. Ontario Intensive Care Study Group, "Evaluation of Right Heart Catheterization in Critically 111 Patients," *Critical Care Medicine* 20(7):928-933, *1992.*

578a. Ontario Ministry of Health, Drug Programs Branch, "Draft National Guidelines for Economic Analysis of Pharmaceutical Products," unpublished draft prepared by A.S. De(sky, Toronto Ontario, May 12, 1992.

579. Opher, D., Research Assistant, Emergency Care Research Institute, Plymouth Meeting, PA, personal communication, September 1993.

580. Orkin, F. K., "Editorial Views: Practice Standards: The Midas Touch or the Emperor's New Clothes'? *Anesthesiology 70:570-571, 1989.*

581. Orleans, M., Professor, Epidemiology and Community Health, Department of Preventive Medicine and Biometrics, School of Medicine, University of Colorado Health Sciences Center, Denver, CO, personal communication, May 25, 1994

582. Outpatient Ophthalmic Surgery Society, "New Federal Government Guidelines Would Reduce Needed Cataract Procedures for Medicare Patients" (press release), San Diego, CA, Feb. 25, 1993.

583. Oxman, A. D., and Guyatt, G. H., "Guidelines for Reading Literature Reviews," *Canadian Medical Association Journal* 138(8):697-703, April 1988.

584. Paradise, J.L, Bluestone, C. D., Bachman, R. Z., et al., "Efficacy of Tonsillectomy for Recurrent Throat Infection in Severely Affected Children: Results of Parallel Randomized and Nonrandomized Clinical Trials," New *England Journal of Medicine* 310:674-683, *1984.*

585. Park, R. E., Fink, A., Brook, R. H., et al., "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures," *American Journal of Public Health* 76(7):766-772,1986.

586. Park, R. E., Fink, A.. Brook, R. H., et al., "Physic ian Rat ings of Appropriate Indications for Three Procedures: Theoretical Indications vs. Indications Used in Practice," *American Journal of Public Health* 79(4):445-447,1989.

587. Parliament, M. B., Danjoux, C. E., and Clayton, T., "IS Cancer Treatment Toxicity Accurately Reported?" *International Journal of Radiation Oncology, Biology, Physics* 11:603-608, *I 985.*

588. Parsonnet, V., "The Proliferation of Cardiac Pacing: Medical, Technical, and Socioeconomic Dilemmas," *Circulation* 65:84*I*-845, *1982.*

589. Patrick, D. L., Bush, J. W., and Chen, M. M., "Methods for Measuring Levels of Well-Being for a Health Status Index," *Health Services Research 8:228-245,* 1973.

590. Patrick, D. L.. and Deyo, R. A., "Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life," *Medical Care* 27(3) (suppl.):S217-232,1989.

591. Patrick, D.L. and Erickson, P., *Health Status and Health Policy: A llocating Resources to Health Care* (New York, NY: Oxford University Press, Inc., I 993).

592. Pauly, M. V., "Effectiveness Research and Impact of Financial Incentives on Outcomes," *Improving Health Policy and Management: Nine Critical Research Issues for the 1990s*, S.M. Shorten and U.E. Reinhart (eds. ) (Ann Arbor, MI: Health Administration Press, 1992).

593. Pearson, K., "Report on Certain Enteric Fever Inoculation Statistics," *British Medical Journal* 2: 1243-1246, 19(M.

594. Pearson, S. D., Margolis, C. Z., Davis, S., et al., "Is Consensus Reproducible'? A Study of an Algorithmic Guidelines Development Process," presented in part at the National Annual Meeting of the Society of Medical Decision Making, Rochester, NY, 1991.

595. Pena, E. F., "Prevention of Abortion," *Americ CC III Journal of Surgery* 87:95-96,1954.

596. Perfetto, E. M., Senior Pharmacoepidemiologist, Center for Medical Effectiveness Research, Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rock~' i lle, MD, personal communication, Aug. 4, 1993, and Feb. 16,1994.

597. Perry, C. E., Houlihan, M., and Greenan, L., "Comparison" of Trends in Normal and Cesarean Section Del iveries," paper presented at annual meeting of the American Public Health Association, Boston, MA, Nov. 13.1988.

598. Perry, S., "Special Report: The Brief Life of the National Center for Health Care Technology," *New England Journal of Medicine* 307: 1095-1100, 1982.

599. Perry, S., "The NIH Consensus Development Program: A Decade Later," *New England Journal of Medicine* 317*(8):485-488,* 1987.

600. Perry, S., and Wilkinson, S. L., "The Technology Assessment and Practice Guidelines Forum: A Modified Group Judgment Method," *International Journal of Technology Assessment in Health Care* 8(2):289-300, *1992.*

601. Peters, W.P., and Rogers. M. C., "Variation in Approval by Insurance Companies of Coverage for Auto] ogous Bone Marrow Transplantation for Breast Cancer," *New England Journal of Medicine* 330(7):473-477, *1994.*

602. Peto, R., "[b]Why Do We Need Systematic Overviews of Random ized Trials'?" *Statistics in Medicine* 6(3):233-240, 1987.

603. Pete, R., "Discussion," *Statistics in Medicine* 6(3):241-*244, 1987.*

604. Peto, R., Collins, R., and Gray, R., "[b]Large-Scale Randomized Evidence: Large. Simple Trials and Overviews of Trials," *Doing More Good than Harm: The Evaluation of Health Care Interventions*, K.S. Warren and F. Mosteller (eds.) (New York, NY: New York Academy of the Sciences, 1993).

605. Phelps, C. E., "The Methodologic Foundations of Studies of the Appropriateness of Medical Care," *New England Journal of Medicine 329(*17): *1241-1245,* 1993.

606. Phelps, C. E., Professor and Chair, Department of Community and Preventive Medicine, and Professor of Political Science and Economics, Un iversity of Rochester, Rochester, NY. "Statement." hearing before the Committee on Finance, Senate. U.S. Congress, Washington, DC, Oct. 26,1993.

607. Physician Payment Review Commission, *Annual Report to Congress 1992* (Washington, DC: U.S. Government Printing Office, 1992).

608. Pierce, Jr., E. C., "The Development of Anesthesia Guidelines and Standards," *Quality Review Bulletin 16:61 -64.* 1990.

609. Pollock, A., "Quality Assurance in Medical Care" (letter), *Journal of Public Health Medicine* 13(3):226-227, **1991.**

**610.** Potosky, A. L., Riley, G. F., Lubitz, J. D., et al., "Potential for Cancer Related Health Services Research Using a Linked Medicare-Tumor Registry Database," *Medical Care 3 1(8):732-748, 1993.*

611. Powe, N. R., Davidoff, A. J., Moore, R. D., et al., "Net Cost from Three Perspectives of Using Low versus High Osmolality Contrast Media in Diagnostic Angiocardiography," *Journal of the American College of Cardiology 21: 1701-1709, 1993.*

612. Powe, N. R., and Griffiths, R. I., "Economic Data and Analysis in Clinical Trials," paper prepared under contract for the Office of Technology Assessment, U.S. Congress, Washington, DC, forthcoming 1994,

613. Power, E. J., Tunis, S. R., and Wagner, J. L., ● 'Technology Assessment and Public Health," *Annual Reviews of Public Health,* 15:561-579, *1994.*

614. "Practice Guidelines for Use of Bone Marrow Transplantation in Breast Cancer," *Health News Daily,* p. *9,* May *12, 1994.*

615. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Making Health Care Decisions: The Ethical and Legal Implications of Informed Consent in the Patient-Practitioner Relationship, Volume Three: Appendices: Studies on the Foundations of Informed Consent* (Washington, DC: U.S. Government Printing Office, October 1982).

616. Prospective Payment Assessment Commission, *Medicare and the American Health Care System: Report to The Congress* (Washington, DC: June 1992)

617. Prospective Payment Assessment Commission, *Report and Recommendations to The Congress* (Washington, DC: March 1992).

618. "Psychologists to Snub AHCPR Depression Guideline Release," *Medical Outcomes and Guidelines Alert 1(7):* 1, 1993.

619. Pugh, D., *The National Institutes of Health: A Bethesda Landmark Celebrates Its Centennial* (Bethesda, MD: National Cancer Institute, National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, 1987).

620. Raiffa, H., *Decision Analysis: Introductory Lectures in Choice Under Uncertainty* (Reading, MA: Addison-Wesley Publishing Company, 1968).

621, Ravnskov, U., "Cholesterol Lowering Trials in Coronary Heart Disease: Frequency of Citation and Outcome," *British Medical Journal 305(6844): 15-19, 1992.*

622. "RAND/AMCC/AMA Project: What Went Wrong?" *Medical Outcomes and Guidelines Alert* 2(5):6, *1994.*

623. Rawles, J., "Castigating QALYs," *Journal of Medical Ethics 15:* 143-147, 1989.

624. Ray, W. A., presentation at "Foresight Seminar on Technology Assessment for Pharmaceuticals in Health Care Reform," The Institute for Alternative Futures, Washington, DC, Dec. 10, 1993.

625. Ray, W. A., Griffin, M. R., and Avorn, J., "Evaluating Drugs After Their Approval for Clinical Use," *New England Journal of Medicine* 329(27):2029-2032, 1993.

626. Ray, W. A., Federspiel, C. F., Schaffner, W., "Prescribing of Tetracycline to Children Less than 8 Years Old: A Two-Year Epidemiologic Study Among Ambulatory Tennessee Medicaid Recipients," *Journal of the American Medical Association 237(* 19):2069-2074, 1977.

627. Ray, W. A., Blazer, D. G., and Schaffer, W., et al., "Reducing Antipsychotic Drug Prescribing for Nursing-Home Patients: A Controlled Trial of the Effect of an Educational Visit," *American Journal of Public Health 77: 1448-1450, 1987.*

628. Reichard, J. and Vibbert, S., (eds.), *The Medical Outcomes and Guidelines Sourcebook* (Washington, DC: Faulkner and Gray, 1993).

629. Redelmeier, D. A., and Tversky, A., '"Discrepancy Between Medical Decisions for Individual Patients and for Groups," *New England Journal of Medicine 322(16): 1162-1164, 1990.*

630. Reesor, N. C., Mamdani, F., Baker, A., et al., 'development and Implementation of Simplified Aminoglycoside Empiric Dosing Guidelines," *Pharmacotherapy* 13:408-414, *1993.*

631. Relman, A. S., "An Institute for Health Care Evaluation," *New England Journal of Medicine 306(* 11):669-670, *1982.*

632. Report on Medical Guidelines & Outcomes Research, "Drug Manufacturers Allot More Resources for Outcomes Research," *Report on Medical Guidelines* 4(14):4, July 15, 1993.

633. Report on Medical Guidelines & Outcomes Research, "NCI Gets Mixed Signals from Advisory Boards on Screening Mammography," *Report on Medical Guidelines & Outcomes Research*, 4(23):5-7, 1993.

634. Retchin, S.M., Fletcher, R.H., Buescher, P.C., et al., "The Application of Official Policy: Prophylaxis Recommendations for Patients With Mitral Valve Prolapse," *Medical Care* 23:1156-1162, 1985.

635. Rettig, R.A., "Health Policy in Radiology: Technology Assessment—An Update," *Investigative Radiology* 26:162-173, 1991.

636. Rice, D., "Estimating the Cost of Illness," unpublished paper, Health Economics Series, No. 6, Public Health Service, U.S. Department of Health Education and Welfare, Washington, DC, 1966.

637. Ridker, P.M., O'Donnell, C.J., Marder, V.J., et al., "A Response to 'Holding GUSTO up to the Light,'" *Annals of Internal Medicine* 120:882-885, 1994.

638. "Report from the Field—HCFA Examines Medicaid Outcomes," *Medical Outcomes & Guidelines Alert* 1(23):9-10, 1993.

639. Riley, G.F., and Lubitz, J., "Longitudinal Patterns in Medicare Costs for Cancer Decedents," *Cancer Care and Cost: DRGs and Beyond*, R.M. Scheffler and N.C. Andrews (eds.) (Ann Arbor, MI: Health Administration Press, 1989).

640. Rizzo, W.B., "Lorenzo's Oil—Hope and Disappointment," *New England Journal of Medicine* 329(11):801-802, 1993.

641. Rizzo, W.B., Leshner, R.T., Odone, A., et al., "Dietary Erucic Acid Therapy for X-Linked Adrenoleukodystrophy," *Neurology* 39:1415-1422, 1989.

642. Robinson, D., and Shettles, L.B., "Use of Diethylstilboestrol in Threatened Abortion," *American Journal of Obstetrics and Gynecology* 63:1330-1333, 1952.

643. Rogers, J.L., Haring, O.M., Wortman, P.M., et al., "Medical Information Systems: Assessing Impact in Areas of Hypertension, Obesity and Renal Disease," *Medical Care* 20:63-74, 1982.

644. Rogers, E. M., *Diffusion of Innovation*, 3rd ed. (London, England: Free Press, 1983).

645. Roos, L. L., and Roos, N. P., "Assessing Existing Technologies: The Manitoba Study of Common Surgical Procedures," *Medical Care* 21:454-462, 1983.

646. Roos, N. P., "Hysterectomy: Variations in Rates Across Small Areas and Across Physicians' Practices," *American Journal of Public Health* 74(4):327-335, 1984.

647. Roos, N. P., and Lyttle, D., "Hip Arthroplasty Surgery in Manitoba: 1973-1978," *Clinical Orthopaedics and Related Research* 199:248-255, 1985.

648. Roos, N. P., and Ramsey, E. W., "A Population-Based Study of Prostatectomy: Outcomes Associated with Differing Surgical Approaches," *Journal of Urology* 137:1184-1188, 1987.

649. Roos, N.P., and Roos, L. L., "Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?" *Medical Care* 20(9):945-958, 1982.

650. Roos, N. P., Wennberg, J. E., Malenka, D. J., et al., "Mortality and Reoperation After Open and Transurethral Resection of the Prostate for Benign Prostatic Hyperplasia," *New England Journal of Medicine* 320:1120-1124, 1989.

651. Roper, W., Winkenwerder, W., Hackbarth, G., et al., "Effectiveness in Health Care: An Initiative to Evaluate and Improve Medical Practice," *New England Journal of Medicine* 319(18):1197-1202.1988.

652. Ross, J. H., President, MetaWorks, Inc., Newton. h4A, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC. Oct. 8, 1993.

653. Ross. J. W., "Further Report on the Use of Diethylstilboestrol in the Treatment of Threatened Abortion," *Journal of the National Medical Association* 45:223, 1953.

654. Rosser, W. W., Hutchison, B. C., McDowell, I., et al., "Use of Reminders to Increase Compliance With Tetanus Booster Vaccination," *Canadian Medical Association Journal* 146:111-917, 1992.

655. Rothman, M.L., Hedrich, S. C., Bulcrot, K. A., et al., "The Validity of Proxy-Generated Scores

as Measures of Patient Health Status," *Medical Care 29(2):* 115-124, 1991.

656. Rothman, M. L., and Revicki, D. A., "Issues in the Measurement of Health Status in Asthma Research," *Medical Care 31(3* suppl.): MS82-96, 1993.

***657.*** Rovira, J., Asociacion de Economia de la Salud, Barcelona, Spain, personal communication, Aug. 27, 1993.

658. Rovira, J. (project leader), '*The Harmonization by Consensus of the Methodology for Economic Evaluation of Health Technologies in the EC," project proposal submitted to the Commission of the European Communities, Brussels, Belgium, unpublished document, undated.

659. Rovira, J., Antoñanzas, F., and Muslera, E., "Study for the Establishment of a System of Evaluation of Health Technologies in Spain: Project 4: Proposal for Standardization of Some Methodological Aspects of Cost-Effectiveness Analyses (CEA), Cost-Utility Analyses (CUA) and Cost-Minimization Analyses (CMA) in the Evaluation of Health Technologies and Programmes for the Purpose of Dec ision Making in Resource Allocation, Fourth Report," Asociacion de Economia de la Salud, Barcelona, Spain, unpublished document, February 1993.

660. Rowe, G., Wright, G., and Bolger, F., "Delphi: A Reevaluation of Research and Theory," *Technological Forecasting and Social Change* 39:235-251, *1991.*

*661.* Royal College of Radiologists Working Party, "Influence of the Royal College of Radiologists' Guidelines on Hospital Practice: A Multicenter Study," *British Medical Journal* 304:740-743, *1992.*

*662.* Rubin, D. B., "A New Perspective," *The Future of Meta-Analysis*, K.W. Wachter and M.L. Straf, (eds.) (New York, NY: Russel Sage Foundation, 1990).

663. Rubin, R., "Cholesterol, Round 2," *U.S. News and World Report,* 114(25)61-66, June 28, 1993.

664. Rubsamen, D. S., "The Standard of Care and Where It Comes From," *Seminars in Anesthesia* 5:237-239, 1986.

***665.*** Runyon, B. A., McHutchison, J. G., Antillon, M. R., et al., "Short-Course versus Long-Course Antibiotic Treatment of Spontaneous Bacterial Peritonitis: A Randomized Controlled Study of 100 Patients," *Gastroenterology* 100: 1737-1742, 1991,

666. Sackett, D. L., "The Cochrane Collaboration" (editorial), *ACP Journal* Club 120(suppl. 3):A-11, May/June 1994.

667. Sackett, D. L., and Torrance, G. W., "The Utility of Different Health States as Perceived by the General Public," *Journal of Chronic Diseases 31*:697-704, *1978.*

*668.* Sacks, H. S., Berrier, J., Reitman, D. et al., "Meta-analyses of Randomized Controlled Trials," *New England Journal of Medicine* 316(8):450-455, *1987.*

*669.* Sacks, H., Chalmers, T., and Smith, H., "Randomized Versus Historical Controls for Clinical Trial s," *American Journal of Medicine* 72:233-240, February 1982.

670. Saver, R., "Reimbursing New Technologies: Why Are the Courts Judging Experimental Medicine?" *Stanford Law Review 44: 1095-*1131, 1992.

671. Schaffner, W., Ray, W. A., Fedspiel, C. F., et al., "Improving Antibiotic Prescribing in Office Practice: A Controlled Trial of Three Educational Methods," *Journal of the Medical Association 250: 1728-1732, 1983.*

*672.* Schatz, G., Attorney, Reed, Smith, Shaw, & McClay, Washington, DC, personal communication, July 13, 1993.

673. Schectman, J. M., Elinsky, E. G., and Pawlson, L. G., "Effect of Education and Feedback on Thyroid Funcation Testing Strategies of Primary Care Clinicians," *Archives of Internal Medicine* 15 1:2 163-2166, 1991.

674. Schoellhorn, R. A., "Federal Technology Assessment Programs Should Be Limited," *Medical Instrumentation* 15(5):291-*292, 1981.*

675. Schoellhorn, R. A., Chairman of the Board, Health Industry Manufacturers Association, Washington, DC, "Statement," *Health Care Technology Reauthorization,* hearing before the Subcommittee on Health and the Environment, Committee on Energy and Commerce, House of Representatives, U.S. Congress, Serial No. 97-27 (Washington, DC: U.S. Government Printing Office, 198 1).

676. Schoenbaum, S. C., and Gottlieb, L, K., "Algorithm Based Improvement of Clinical Quality," *British Medical Journal 301:* 1374-1376, 1990.

677. Schoenbaum, S. C., and Murrey, K. O., "Impact of Profiles on Medical Practice," *Conference on Profiling*, No. 92-2, Physician Review Payment Commission (cd. ) (Washington, DC: PPRC, May 1992).

678. Schriger, D., UCLA Emergene> Medicine Center, UCLA School of Medicine, Los Angeles, CA, personal communication, June 7, 1994.

679. Schroeder, S. A., "Strategies for Reducing Medical Costs by Changing Physicians' Behavior: Efficacy and Impact on Quality of Care," *International Journal of Technology Assessment in Health Care* 3: 39-50, 1987.

680. Schuyler, P., presentation at "An Evidence-Based Health Care System: The Case for Clinical Trials Registries, " conference sponsored by National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, Dec. 6-7, 1993.

681. Schwartz. J. S., "Decision Support Systems and Their Potential Contribution to Consensus Development," paper presented at "Workshop to Improve Group Judgment for Medical Practice and Technology Assessment," sponsored by the Council on Health Care Technology, Division of Health Care Services, Institute of Medicine, Washington, DC, May 15-16, 1990.

682. Schwartz, J. S., and Cohen, S. J., "Changing Physician Behavior," *Conference Proceedings. Primary Care Research: An Agenda for the 90s*, J. Mayfield and M. I.. Grady (eds.) (Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, September 1990).

683. Scott, E. A.. and Black, A., "When Does Consensus Exist in Expert Panels'?" *Journal of Public Health Medicine* I 3(1):35-39, 1991.

684. Scott, E., and Black, N., "When Does a Consensus Exist in Expert Panels?" (letter), *Journal of Public Health Medicine* 13(4):343-4, 1991.

685. Senekjian, E. K., Potkul, R. K., Frey, K., et al., "Infertility Among Daughters Either Exposed or Not Exposed to Diethylstilboestrol '" *American Journal of Obstetrics and Gynecology* 1 58:493-498, 1988.

686. Shekelle, P. G., RAND Corp., Santa Monica, CA, personal communication, Aug. 1, 1994.

687. Shekelle, P. G., Adams, A. H.. Chassin, M. R., et al., "Spinal Manipulation for Low-Back Pain," *Annals of Internal Medicine* 117(7):590-598, 1992.

688. Shekelle, P. G., Brook, R. H., "A Community Based Study of the Use of Chiropractic Services," *American Journal of Public Health* 81(4):439-442, April 1991.

689. Shekelle, P. G., and Schriger, D. L.. *Using the Appropriateness Method in the Agency for Health Care Policy and Research: Clinical Practice Guideline Development*, RAND MRR 233-AHCPR (Santa Monica, CA: RAND cm-p., 1 993).

690. Sherman, C. R., Potosky, A.L., Weis, K. A.. et al., "The Consensus Development Program: Detecting Changes in Medical Practice Following a Consensus Conference on the Treatment of Prostate Cancer," *International Journal of Technology Assessment in Health Care* 8;683-693, 1992.

691. Sherrington, A. M.. An Annotated Bibliography of Studies on the Flow of Medical Information to Practitioners, *Methods of Information in Medicine* 4:45-57, 1965.

692. Shortel I, S.M. and Reinhart, U.E. (eds. ), *Improving Health Policy and Management: Nine Critical Research Issues for the 1990s* (Ann Arbor, MI: Health Administration Press, 1992).

693. Showstack, J. A., Rosenfeld, K. E., Car-nick, D. W., et al., "Association of Volume With Outcome of Coronary Artery Bypass Graft Surgery," *Journal of the American Medical Association* 257(6):785-789, 1987.

694. Showstack, J., Katz, P., Amend, W'.. et al., "The Association of Cyclosporine with the 1 -Year Costs of Cadaver-Donor Kidney Transplants," *Journal of the American Medical Association* 264(14): 1818-1823, 1990.

695. Shulkin, D. J., Fox, K. R., and Stadtmauer, E. A., "Guidelines for Prophylactic Platelet Transfusions: Need for a Concurrent Outcomes Management System," *Quality Review Bulletin* 18( 12):477-479, December 1992.

696. Sickles, E. A., and Kopans, D. B.. "Deficiencies in the Analysis of Breast Cancer Screening Data" (editorial), *Journal of the National Cancer Institute* 85(20): 1621-1624. October 1993.

697. Siegal, J., remarks at fifth meeting of the Cost-Effectiveness Panel on Clinical Preventive Services. Office of Disease Prevention and Health Promotion, Office of the Assistant Director for

Health, Public Health Service, U. S. Department of Health and Human Services, Washington DC, Mar. 8-9, 1994.

698. Silk-man, C.E., "From the Patient's Bed," *Health Management Quarterly 13(2): 12-15, 1991.*

699. Silberman, C. E., New York, N}'. personal communication, Apr. 19. 1994.

700. Silberman, G., Assistant Director, Program Evaluation Methodology Division, [1. S. General Accounting Office, Washington, DC. personal communication, July 2, 1992.

701. "Sleeping Giant" (editorial), *Lancet* 337 (8749): 1067, 1991.

702. Simon, G. E., Daniell, W., Stockbridge, H., et al., "Immunologic, Psychological and Neuropsycholog ical Factors i n Mult iple Chem ical Sensitivity: A Controlled Study," *Annals of Internal Medicine* 19(2):97-103, 1993.

703. Sims, P. D., Cabral, D., Daley, W., et al., "The Incentive Plan: An Approach to Modification of Physician Behavior," *American Journal of Public Health 74(2): 150-152, 1* 984.

704. Simmer, T. L., Nerenz, D.R., Rutt, W.M., et :11., "A Randomized, Controlled Trial of an Attending Staff Serv ice in General Internal Medicine," *Medical Care* 29(7):JS3 *1-40, 1* 99 *1.*

705. Sisk, J. E., "Improving the Use of Research-Based Evidence in Policy Making: k-, ffective Care in Pregnancy and Childbirth in the United States," *Milbank Quarterly* 71 (3):477-496, 1993.

706. Sloan, F., Morrisey, M., and Valvena, J., "Medicare Prospective Payment and the Use of Medical Technolog ies in Hospitals," *Medic 111 Care* 21 (9):837-853, 1988.

707. Smith, R. P., Meier, P., Abrams, D., et al., "Time for Some Really Large, Simple Randomized Trials in HIV Disease," *Oncology* 6:63-64, 1992.

707a. Smith, G. V., and O.W. Smith, "Prophylactic hormone therapy. Relation to Complications of Pregnancy," *Obstetrics and Gynecology* 4: 129-141, 1954.

708. Smits, H. L., "Payer Considerations: Experience of the Medicare Program," presentation at "Cost-Effectiveness Analysis: Assessing the Value of Medical Technology," conference sponsored by Health Care Technology Institute, Washington, DC, Apr. 12, 1994.

709. Snider, D., Acting Deputy Associate Director for Science, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, April 1994.

710. Soumerai, S.B., and Avorn, J., "Efficacy and Cost-Containment in Hospital Pharmacotherapy: State of the Art and Future Directions," *Milbank Quarterly* (Health and Society) 62:447-474, 1984.

711. Soumerai, S.B., and Avorn, J., "Economic and Policy Analysis of University-Based Drug 'Detailing'," *Medical Care* 24(4):313-331, 1986.

712. Soumerai, S.B., and Avorn, J., "Principles of Educational Outreach (Academic Detailing) to Improve Clinical Decision Making," *Journal of the American Medical Association* 263:549-556, 1990.

713. Soumerai, S.B., McLaughlin, T.J., and Avorn, J., "Improving Drug Prescribing in Primary Care: A Critical Analysis of the Experimental Literature," *Milbank Quarterly* 67(2):268-317, 1989.

714. Soumerai, S.B., Ross-Degnan, D., Avorn, J., et al., "Effects of Medicaid Drug-Payment Limits on Admission to Hospitals and Nursing Homes, *New England Journal of Medicine* 325:1072-1077, 1991.

715. Soumerai, S.B., Ross-Degnan, D., Gortmaker, S., et al, "Withdrawing Payment for Nonscientific Drug Therapy," *Journal of the American Medical Association* 263(6):831-839, 1990.

716. Soumerai, S.B., Salem-Schatz, S., Avorn, J., et al., "A Controlled Trial of Educational Outreach to Improve Blood Transfusion Practice," *Journal of the American Medical Association* 270(8):961-966, 1993.

717. Sox, H.C., and Woolf, S.H., "Evidence-Based Practice Guidelines from the US Preventive Services Task Force" (editorial), *Journal of the American Medical Association* 269(20):2678, 1993.

718. Speigel, J.S., Shapiro, M.F., Berman, B., et al., "Changing Physician Test Ordering in a Uni-

versity Hospital: An Intervention of Physician Participation, Explicit Criteria, and Feedback," *Archives of Internal Medicine* 149:549-553, *1989.*

*719.* Stafford, R., "Alternative Strategies for Controlling Rising Cesarean Section Rates," *Journal of the American Medical Association* 263(5):683-687, *1990.*

*720.* Stason, W., and Weinstein, M. C., 'allocation of Resources to Manage Hypertension," *New England Journal of Medicine 296( 13):732-739,* 1977.

*721.* State of New York, Department of Health, "Backgrounder on Cardiac Surgeon Data," Albany, NY, Dec. 18, 1991, reprinted in S. Vibbert and J. Reichard (eds. ), The *Medical Outcomes and Guidelines Sourcebook, 1992 Edition* (Washington, DC: Faulkner and Gray, 1992).

722. State of Oregon, "Waiver Application, Oregon Medicaid Demonstration Project," submitted to U.S. Department of Health and Human Services, Health Care Financing Administration, Baltimore, MD, August 1991.

723. State of Minnesota, Department of Health, Haugen, D, Minneapolis, MN, personal communication, March 1992.

724. Steinberg, E. P., Professor, Division of Internal Medicine, Johns Hopkins University, Baltimore, MD, personal communication, May 20, 1994.

725. Stewart, A., Greenfield, S., Hays, R. D., et al., "Functional Status and Well-Being of Patients with Chronic Conditions: Results from the Medical Outcomes Study," *Journal of the American Medical Association 262(7):907-913, 1989.*

*726.* Stewart, A., and Ware, J. (eds.), *Measuring Functioning and Well-Be Being: The Medical Outcomes Study Approach* (Durham, NC: Duke University Press, 1992).

727. Stiell, I. G., McKnight, R. D., Greenberg, G. H., et al., *Journal of the American Medical Association 271 (11 ):*827-832, 1994.

728. Stillman, R. J., "In Utero Exposure to Diethylstilbestrol: Adverse Effects on the Reproductive Tract and Reproductive Performance in Male and Female Offspring," *American Journal of Obstetrics and Gynecology* 142:905-92 *1, 1982*

729. Stillman, L., Data-Base Coordinator, Vermont-Oxford Trials Facilitation Service, Burlington, VT, personal communication, Mar. 7, 1994.

730. Stirrat, G.M., Farndon, J., Farrow, S.C., et al., "The Challenge of Evaluating Surgical Procedures," *Annals of the Royal College of Surgeons of England* 74:80-84, 1992.

731. Stocking, B., *Initiative and Inertia: Case Studies in the NHS* (London, England: Nuffield Provincial Hospitals, 1985).

732. Stone, V.E., Seague, III, G.R., Hertz, T., et al., "The Relation Between Hospital Experience and Mortality for Patients With AIDS," *Journal of the American Medical Association* 268(19) 2655-2661, 1992.

733. Strauss, M.J., Gong, J., Gary, B.D., et al., "The Cost of Home Air-Fluidized Therapy for Pressure Sores: A Randomized Controlled Trial," *Journal of Family Practice* 33(1):52-59, 1991.

734. Street, D.A., and Javitt, J.C., "National Five-Year Mortality After Inpatient Cataract Abstraction," *American Journal of Ophthalmology* 113:263-268, 1992.

735. Strosberg, M.A., Weiner, J.M., Baker, R., et al., (eds.) *Rationing America's Medical Care: The Oregon Plan and Beyond* (Washington, DC: The Brookings Institution, 1992).

736. Stross, J.K., and Bole, G.G., "Evaluation of a Continuing Education Program in Rheumatoid Arthritis," *Arthritis and Rheumatism* 23(7): 846-849, 1980.

737. Stross, J.K. and Harlan, W.R., "The Dissemination of New Medical Information," *Journal of the American Medical Association* 241: 2622-2624, 1979.

738. Stuart, M.E., Hanley, M.A., Thompson, R.S., et al., "Clinical Practice and New Technology: Prostate-Specific Antigen (PSA)," *HMO Practice* 6:5-11, 1993.

739. Siu, A., Deputy Commissioner, New York State Department of Health, Albany, NY, personal communication, Oct. 12, 1993.

740. Siu, A.L., Sonnenberg, F.A., Manning, W.G., et al., "Inappropriate Use of Hospitals in a Randomized Trial of Health Insurance Plans," *New England Journal of Medicine* 315(20):1259-1266, 1986.

741. Sussman, E.J., Tsiaras, W.G., Soper, K.A., "Diagnosis of Diabetic Eye Disease," *Journal*

*of the American Medical Association 247: 3231-3234,* 1982.

*742.* Sutherland, H. J., Dunn, V., and Boyd, N. F., "Measurement of Values for States of Health With Linear Analog Scale s," *Medical Decision Making* 3:477-487, *1983.*

*743.* Swedlow, A., Johnson, G., Smithline, N., et al., "Increased Costs and Rates of Use in the California Workers' Compensation System as a Result of Self-Referral by Physicians," *New England Journal of Medicine 327(21): 1502-1506, 1992.*

*744.* Taffel, S. M., and Placek, P, J., "An Overview of Recent Patterns in Cesarean Delivery and Where We Stand Today," paper presented at annual meeting of the American Public Health Association, Boston, MA, Nov. 15, 1988.

745. Tarlov, A. R., Director, Health Institute, New England Medical Center, Boston, MA, personal communication, July )5, 1993.

746. Tarlov, A. R., Ware, J. E., Greenfield, S., et al., "The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care," *Journal oj'the American Medical Association* 262(7):925-930, 1989.

*747.* Temple, R., "Problems in the Use of Large Data Sets to Assess Effectiveness," *International Journal of Technology Assessment in Health Care* 6(2):211-219, 1990.

748. Temple, R., Director, Office of Drug Evaluation, Center for Drug Evaluation and Research, Food and Drug Administration, U.S. Department of Health and Human Services, Rockville, MD, personal communication, February 1994.

749. Teutsch, S. M., "A Framework for Assessing the Effectiveness of Disease and Injury Prevention," *Morbidity and Mortality Weekly Report 41( RR-3),* Mar. 27, 1992.

750. Teutsch, S. M., presentation at third meeting of the Cost-Effectiveness Panel on Clinical Preventive Services, Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Public Health Service, U.S. Department of Health and Human Services, Washington, DC, Sept. 28, 1993.

751. Teutsch, S. M., Chief, Prevention Effectiveness Activity, Epidemiology Program Office, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 22, 1994.

752. Teutsch, S. M., Chief, Prevention Effectiveness Activity, Epidemiology Program Office, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA, personal communication, Jan. 27, 1994, and July 5, 1994.

753. Teutsch, S. M., Luce, B. R., and Lewder, D. O., "CDC Hones in on Prevention," *HMO Magazine* 34(4):47-49, 1993.

754. Thacker, S. B., Koplan, J. P., Taylor, W. R., et al., "Assessing Prevention Effectiveness Using Data to Drive Program Decisions," *Public Health Reports 109(2): 187-194, 1994.*

*755.* Thompson, L. W., Executive Vice President, Eli Lilly & Company, Indianapolis, IN, personal communication, July 22, 1993.

756. Thorpe, K. E., "Health Care Cost Containment: Results and Lessons From the Past 20 Years," *Improving Health Policy and Management: Nine Critical Research Issues for the 1990s,* S.M. Shorten and U.E. Reinhart (eds.) (Ann Arbor, MI: Health Administration Press, 1992).

757. Thum, A., presentation at "An Evidence-Based Health Care System: The Case for Clinical Trials Registries," conference sponsored by National Institutes of Health, Public Health Service, U.S. Department of Health and Human Services, Bethesda, MD, Dec. 6-7, 1993.

758. Tiemey, W. M., Hui, S. L., and McDonald, C.J., "Reminders to Physicians from an Introspective Computer Medical Record," *Annals of Internal Medicine* 100(1): 130-138, 1984.

759. Tierney, W. M., Hui, S. L., and McDonald, C. J., "Delayed Feedback of Physician Performance Versus Immediate Reminders to Perform Preventive Care: Effects on Physician Compliance," *Medical Care* 24(8):659-666, *1986.*

*760.* Tiemey, W, M., Miller, M. E., and McDonald, C. J., "The Effect on Test Ordering of Informing Physicians of the Charges for Outpatient Diagnostic Tests," *New England Journal of Medicine* 322: 1499-1504, 1990.

761. Tiemey, W. M., Miller, M. E., Hui, S. L., et al., "Practice Randomization and Clinical Research: The Indiana Research," *Medical Care 29(7) :JS57-JS64, 1991.*

762. Torrance, G. W., 'measurement of Health State Utilities for Economic Appraisal-A Review," *Journal of Health Economics 5( 1 ): 1-30,* 1986.

*763.* Torrance, G., "Utility Theory, QALYs, and Measurement Instruments: Uses for Policy," presentation at inter-Depm-tmental Discussion Group on CEA/Clinical Preventive Services, Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Public Health Service, U.S. Department of Health and Human Services, Washington DC, Dec. 11, 1992.

764. Total Knee Replacement Patient Outcomes Research Team, Indiana University, Indianapolis, IN, "Continuation Application," submitted to the Agency for Health Care Policy and Research, Pub] ic Health Service, U.S. Department of Health and Human Services, Rockville, MD, December 1993.

765. Towers Perrin, "Emerging Technology Survey Results," Rosslyn, VA, unpublished document, October 1993.

766. Trontell, A., Office of Research and Demonstrations, Health Care Financing Administration, U.S. Department of Health and Human Services, Baltimore, MD. personal communication, Dec. 14, 1993.

767. Tunis, S. R., Bass, E. B., and Steinberg, E. P., "The Use of Angioplasty, Bypass Surgery, and Amputation in the Management of Peripheral Vascular Disease," *New England Journal of Medicine* 325:556-562, 1991.

768. Tunis, S. R., Hayward, R. S. A., Wilson, M. C., et al., "Internists' Attitudes about Clinical Practice Guide lines," *Annuls of Internal Medicine* 120:956-963, 1994.

769. Turner, J. A., Ersek, M., Herron, L., "Patient Outcomes After Lumber Spinal Fusions," *Journal of the American Medical Association* 268(7):907-911, 1992.

770. Turner, J. A., Ersek, M., Herron, L. et al., "Surgery for Lumbar Spinal Stenosis: Attempted Meta-Analysis of the Literature," *Spine* 17( 1): I -8, 1992.

771. Udvarhelyi, 1. S., Colditz, G. A., Rai, A., et al., "Cost-Effectiveness and Cost-Benefit Analyses in the Medical Literature, " *Annals of Internal Medicine 11* 6(3 ):238-244, 1992.

772. Udvarhelyi, 1. S.. Gatsonis, C., Epstein, A. M., et al.. "Acute Myocardial infarction in the Medicare Population: Process of Care and Clinical Outcome s." *Journal of the American Medical Association 268(* 18):2530-2536, 1992.

773. Uhlmann, R. F., Pearlman, R. A., and Cain, K. C., "'Physicians' and Spouses' Predictions of Elderly Patients' Resuscitation Preferences," *Journal of Gerontology* 43(5) :M115-M 121, 1988.

774. U.S. Congress, House of Representatives, Committee on Science and Astronautics. *Office of Technology Assessment: Background and Status*, prepared by the Congressional Research Service, Comm. Print No. 99-672 (Washington, DC: U.S. Government Printing Office, 1973).

775. U.S. Congress, House of Representatives, Committee on Energy and Commerce, Subcommittee on Health and the Environment, *Health* and *the Environment Miscellaneous — Part 4*, hearing. Apr. 24 and 27. 1984, Serial No. 98-141 (Washington, DC: U.S. Government Printing Office. 1984).

776. U.S. Congress, House of Representatives, Committee on Energy and Commerce, Subcommittee on Health and the Environment, *Deficit Reduction Proposals Affecting Medi - (arc,* hearing, July 16, 1986, Serial No. 99-161 (Washington, DC: U.S. Government Printing Office. 1 987).

777. U.S. Congress, Library of Congress, Congressional Research Service, "Breast Cancer: Selected References, 1990- 1993" prepared by B.F. Mangan, Washington, DC. June 1993.

778. U.S. Congress, Office of Technology Assessment. *Development of Medical Technology: Opportunities for Assessment*, OTA-H-34 (Washington, DC: U.S. Government Printing Office, August 1976).

779. U.S. Congress, Office of Technology Assessment, *Assessing tile Efficacy and Safety of Medical Technologies*, OTA-H-75 (Washington, DC: U.S. Government Printing Office, September 1978).

780. U.S. Congress, Office of Technology Assessment. *The Implications of Cost-Effectiveness Analysis of Medical Technology*, OTA-H - 126 (Washington. DC: U.S. Government Printing Office, 1980).

781. U.S. Congress, Office of Technology Assessment, *The Implications of Cost Effectiveness Analysis of Medical Technology—Background*

*Paper #1: Methodological Is.sues and Literature Review*, OTA-BP-H-5 (Washington, DC: U.S. Government Printing Office, September 1980).

782. U.S. Congress, Office of Technology Assessment, *Policy Implications of the Computed Tomography (CT) Scanner: An Update* (Background Paper to Update OTA'S 1978 Report entitled *Policy Implications of the Computed Tomography (CT) Scanner),* OTA-BP-H-8 (Washington, DC: U.S. Government Printing Office, January 1981 ).

783. U.S. Congress, Office of Technology Assessment. *Strategies for Medical Technology Assessment*, OTA-H-181 (Washington, DC: U.S. Government Printing Office. September 1982).

784. U.S. Congress, Office of Technology Assessment, *The Impact of Randomized Clinical Trials on Health Policy and Medical Practice—Background Paper*, OTA-BP-H-22 (Washington, DC: U.S. Government Printing Office, August 1983).

784a. U.S. Congress, Office of Technology Assessment, *Federal Policies and the Medical Devices Industry*, OTA-H-229 (Washington, DC: U.S. Government Printing Office, October 1984).

785. U.S. Congress, Office of Technology Assessment, *Third Report on the Prospective Payment Commission (ProPAC)* (Washington, DC: U.S. Government Publishing Office, November 1 987).

786. U.S. Congress, Office of Technology Assessment, *The Quality of Medical Care: Info - tion for Consumers*, OTA-H-386 (Washington, DC: U.S. Government Printing Office, June 1988).

786a. U.S. Congress. Office of Technology Assessment, *Screening for Open Angle Glaucoma in the Elderly* (Washington, DC: U.S. Government Printing Office, October 1988).

787. U.S. Congress, Office of Technology Assessment, *Unconventional Cancer Treatments*, OTA-H-405 (Washington, DC: U.S. Government Printing Office, September 1990).

788. U.S. Congress, Office of Technology Assessment, *Evaluation of the Oregon Medicaid Proposal, OTA-H-531* (Washington, DC: U.S. Government Printing Office, May 1992).

789, U.S. Congress, Office of Technology Assessment, *Pharmaceutical R&D: Costs, Risks and Rewards*, OTA-H-522 (Washington, DC: U.S. Government Printing Office, February 1993).

790. U.S. Congress, Office of Technology Assessment, *Defensive Medicine and Medical Malpractice OTA-H-602* (Washington, DC: U.S. Government Printing Office, July 1994).

791. U.S. Congress, Senate, Special Committee on Aging, *Fraud, Waste and Abuse in the Medicare Pacemaker Industry* (Washington, DC: U.S. Government Printing Office, 1982).

791 a. U.S. Congress, Senate, Committee on Appropriations, Subcommittee on Health and Labor, Health and Human Services, Education and Related Agencies, *Variations in Medical Practice,* special hearing, Nov. 19, 1984, Serial No. 98-1239 (Washington, DC: U.S. Government Printing Office, 1985).

792. U.S. Congress, Senate, Committee on Finance, Subcommittee on Health, *Medicare Patient Outcome Assessment Research,* hearing, July 11, 1988, Serial No. 100-960 (Washington, DC: U.S. Government Printing Office, 1989).

793. U.S. Department of Health and Human Services, Health Care Financing Administration, *Medicare Coverage Manual,* HCFA Pub. 6, section 65-6, information provided by Ron Peyton, Baltimore, MD, June 1994.

794. U.S. Department of Health and Human Services, Health Care Financing Administration, Washington, DC, letter (approving Oregon waiver proposal, grant no. 11 -P-90160/O-01) to Oregon Department of Human Resources, Salem, OR, Mar. 19, 1993.

795. U.S. Department of Health and Human Services, Health Care Financing Administration, Office of Research and Demonstrations, *Health Care Financing Program Statistics: Medicare and Medicaid Data Book, 1990,* HCFA Pub. No. 03314 (Washington, DC: U.S. Government Printing Office, March 1991 ).

796. U.S. Department of Health and Human Services, Health Care Financing Administration, Office of Research and Demonstrations, *Health Care Financing Status Report: Research and Demonstrations in Health Cure Financing, Fiscal Year 1992 Edition,* HCFA Pub. No. 03336 (Baltimore, MD: January 1993).

797. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "Medical Treatment Effectiveness Research," *AHCPR Program Note*, (Rockville, MD: March 1990).

798. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. *AHCPR: Purpose and Programs*, Pub. No. OM-O096, (Rockville, MD: September 1990).

799. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Report to Congress: The Feasibility of Linking Research-Related Data Bases to Federal and Non-Federal Medical Administrative Data Bases*, AHCPR Pub. No. 91-0003 (Washington, DC: U.S. Government Printing Office, 1991).

800. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. "Medical Treatment Effectiveness Research: A View from Inside the Agency for Health Care Policy and Research," *AHCPR Program Note* Pub. No. 91-0025, (Rockville, MD: June 1991).

801. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Acute Pain Management: Operative Medical Procedures and Trauma*, Clinical Practice Guideline, prepared by D.B. Carr, A.K. Jacox, C.R. Chapman, et al., AHCPR pub. No. 92-0032 (Rockville, MD: February 1992).

802. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Urinary Incontinence in Adults, Clinical* Practice Guideline, prepared by K.A. McCormick, A. Diokno, J. C'oiling, et al., AHCPR Pub. No. 92-0038 (Rockville, MD: March 1992).

803. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Pressure Ulcers in Adults: Prediction and Prevention*, Clinical Practice Guideline No. 3, prepared by N.C. Bergstrom, R.M. Allman, C.E. Carlson, et al., AHCPR Pub. No. 92-0047 (Rockville, MD: May 1992).

804. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "Effective Dissemination of Health and Clinical Information and Research Findings," *AHCPR Grant Announcement*, (Rockville, MD: March 1992).

805. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "Dissemination-Related Research Portfolio," internal memorandum, December 1992.

806. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Cataract in Adults: Management of Functional Impairment*, Clinical Practice Guideline No. 4, prepared by D.M. O'Day, A.J. Adams, E.H. Cassem, et al., AHCPR Pub. No. 93-0542 (Rockville, MD: February 1993).

807. U.S. Department of Health and Human Services. Public Health Service, Agency for Health Care Policy and Research, workshop on "Medical Effectiveness Research: Strategies for the Future" (comments and discussion of participants at the workshop), Washington, DC, Feb. 17-18, 1993.

808. U.S. Department of Health and Human Services. Public Health Service, Agency for Health Care Policy and Research, *Depression in Primary Care, Vol. 1: Detection and Diagnosis*, Clinical Practice Guideline No. 5, prepared by A.J. Rush, W.E. Golden, G.W. Hall, et al., AHCPR Pub. No. 93-0550 (Rockville, MD: April 1993).

809. U.S. Department of Health and Human Services. Public Health Service, Agency for Health Care Policy and Research, *Depression in Primary Care, Vol. 2: Treatment of Major Depression*, Clinical Practice Guideline No. 5, prepared by A.J. Rush, W.E. Golden, G.W. Hall, et al., AHCPR Pub. No. 93-0551 (Rockville, MD: April 1993).

810. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Sickle Cell Disease: Screening, Diagnosis, Management, and Counseling in Newborns and Infants*, Clinical Practice Guideline No. 6, prepared by J.E. Smith, T.R. Kinney, B. Ames. et al., AHCPR Pub. No. 93-0562 (Rockville, MD: April 1993).

811. U.S. Department of Health and Human Services, Public Health Service, Agency for

Health Care Policy and Research, "Medical Treatment Effectiveness Research: PORT-II," AHCPR grant announcement, August 1993.

812. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "Clinical Practice Guideline Development," *AHCPR Program Note,* AHCPR Pub. No. 93-0023 (Rockville, MD: August 1993).
813. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "AHCPR Evaluation of Clinical Practice Guidelines," unpublished draft, Rockville, MD, September 1993.
814. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Fiscal Year 1994, Volume IX, *Justification of Estimates for Appropriations Committees* (Washington, DC: U.S. DHHS, 1993).
815. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Office of the Forum for Quality and Effectiveness in Health Care, workshop on "Levels of Evidence," Rockville, MD, Nov. 1-2, 1993.
816. U.S. Department of Health and Human Services, Public Health Service. Agency for Health Care Policy and Research, meeting of the National Advisory Council for Health Care Policy, Research, and Evaluation, Bethesda MD, May 27, 1993.
817. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, "Medical Treatment Effectiveness Program (MEDTEP) Research Projects," unpublished documents, Rockville, MD, continuously updated.
818. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Evaluation and Management of Early HIV Infection*, Clinical Practice Guideline No. 7, prepared by W. El-Sadr, J.M. Oleske, B.D. Agins, et al., AHCPR Pub. No. 94-0572 (Rockville, MD: January 1994).
819. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. *Benign Prostatic Hyperplasia: Diagnosis and Treatment, Clinical* Practice Guideline No. 8, prepared by J.D. McConnell, M.J. Barry, R.C. Bruskewitz, et al., AHCPR Pub. No. 94-0582 (Rockville, MD: U.S. DHHS, February 1994).
820. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, *Management of Cancer Pain*. Clinical Practice Guideline No. 9. prepared by A. Jacox, D.B. Carr, R. Payne, et al.. AHCPR Pub. No. 94-0592 (Rockville, MD: U.S. DHHS, March 1994).
821. U.S. Department of Healthand Human Services, Public Health Service, Agency for Health Care Policy and Research, Rockville, MD, letter and memoranda to Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 22, 1994.
822. U. S. Department of Health and Human Services, Public Health Sel%'ice, Agency for Health Care Policy and Research, Request for Application. "Evaluation of the Use of Clinical Practice Guide lines," RFA No. HS-94-007, *NIH Guide* 23(*16),* Apr. 29, 1994.

822a. U.S. Department of Health and Human Ser- \ ices, Public Health Service, Agency for Health Care Policy and Research, *Quick Reference Guide for Clinicians: Unstable Angina,* AHCPR Pub. No. 94-0603 (Rockville, MD: March 199-l).

823. U.S. Department of Health and Human Services, Public Health Serv ice, Agency for Health Care Policy and Research, "The Rand/UCLA Appropriateness Method," prepared by R.H. Brook, in press.
824. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Center for Medical Effectiveness Research, Rockville, MD, staff personal communication, February 1993.
825. U.S. Department of Health and Human Services, Public Health Serv ice, Agency for Health Care Policy and Research, Office of Health Technology Assessment, *Health Technology Review — Procuren: A Platelet-Derived Wound Healing Formula*, prepared by

S.S. Hotta, and T.V. Holohan (Rockville, MD: July 1992).

826. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Office of Health Technology Assessment, Rockville, MD, "Assessments and Reviews 1981- ," unpublished document, Rockville, MD, 1993.

827. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Office of Health Technology Assessment, meeting of the National Advisory Council for Health Care Policy, Research, and Evaluation, Technology Assessment Task Force, workshop materials, Arlington, VA, Sept. 22, 1993.

828. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "CDC: Forty Years of Prevention," *Dateline: CDC* 18(5), June 1986.

829. U.S. Department of Health and Human Resources, Public Health Service, Centers for Disease Control and Prevention, "Prevention and Control of Influenza: Part I, Vaccines," and "Recommendations of the Advisory Committee on Immunization Practices (ACIP)," *Morbidity and Mortality Weekly Report* 43(RR-9):1-13, May 27, 1994.

830. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "Programs for the Prevention of Suicide Among Adolescents and Young Adults" *Morbidity and Mortality Weekly Report* 43(RR-6):3-18, Dec. 18, 1992.

831. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "Assessing the Effectiveness of Prevention: A Practical Guide to Decision and Economic Analysis," unpublished draft, Atlanta, GA, April 1993.

832. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "Update: Hantavirus Disease—Southwestern United States, 1993," *Morbidity and Mortality Weekly Report* 42:570-572, 1993.

833. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "Recommendations of the Advisory Committee on Immunization Practices (ACIP): Use of Vaccines and Immune Globulins in Persons with Altered Immunocompetence," *Morbidity and Mortality Weekly Report* 42( RR-4), 1993.

834. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "Recommendations for the Prevention and Management of *Chlamydia trachomatis* Infections, 1993," *Morbidity and Mortality Weekly Report 42(RR-12):1-39*, Aug. 6, 1993.

835. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "*Haemophilus b* Conjugate Vaccines and a Combined Diphtheria, Tetanus, Pertussis, and *Haemophilus b* Vaccine" *Morbidity and Mortality Weekly Report* 42(RR-*13):1-15.* Sept. 17, 1993.

836. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, "1993 Sexually Transmitted Diseases Treatment Guidelines," *Morbidity and Mortality Weekly Repot-t 42(RR-14)*, Sept. 24, 1993.

836a. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, letter to Office of Technology Assessment, Washington, DC, July 21, 1993,

837. U.S. Department of Health and Human Services. Public Health Service, Centers for Disease Control and Prevention, "General Recommendations on Immunization," *Morbidity and Mortality Weekly Report* 43(RR-1 ), 1994.

*838.* U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, Rockville, MD, contract with the American Gastroenterological Association, Bethesda, MD, Contract No. 282-94-2023, May 9, 1994.

839. U.S. Department of Health and Human Services, Public Health Service, Food and Drug Administration, FDA staff, discussion with the Office of Technology)' Assessment, U.S. Congress, Washington, DC, April 1993.

839:1. U.S. Department of Health and Human Services, Public Health Service, National Center for Health Services Research. "Program Note," 1988.

*840.* U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, "Origins of the National Institutes of Health," Bethesda, MD, undated.

841. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *Cesarean Childbirth: Report of the NICHD Task Force on Cesarean Childbirth,* NIH Pub. No. 82-2067 (Bethesda, MD: 1981).

842. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *Cost Savings Resulting from NIH Research Support. A Periodic Evaluation of the Cost-Benefits of Biomedical Research,* NIH Pub, No. 90-3109 (Bethesda, MD: National Institutes of Health, October 1990).

843. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *NIH Data Book 1992* (Bethesda, MD: September 1992).

844. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *N/H Data Book 1993,* NIH Pub. No. 93-1261 (Bethesda, MD: September 1993).

845. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *Investment for Humanity: A Strategic Vision for the National Institutes of Health* (Rockville, MD: 1993).

846. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, letters to the Office of Technology Assessment, Washington, DC, responding to survey on relevant effectiveness research activities from the following institutes: National Heart, Lung, and Blood Institute, National Institute for Allergies and Infectious Disease, National Institute of Child Health and Human Development, National Institute of Dental Research, National Eye Institute, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Mental Health, 1993.

847. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, *Quality of Life Assessment: Practice, Problems, and Promise—Proceedings of a Workshop, Oct. 15-17, 1990,* NIH Pub. No. 93-3503, (Bethesda, MD: 1990).

*848.* U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, *Quality of Life Assessment in Cancer Clinical Trials: Report of the Workshop on Quality of Life Research in Cancer Clinical Trials, June* /6-/7, 1990 (Bethesda, MD: 1990).

849. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute staff and other National Institutes of Health staff discussions and materials provided to OTA, 1993.

850. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, *Breast Cancer: A National Strategy: A Report to the Nation,* October 1993.

851. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, Division of Cancer Treatment, and Georgetown University Medical Center, Clinical Economics Research Unit, Lombardi Comprehensive Cancer Center, Institute for Health Care Research and Policy with the Office of Continuing Medical Education, *NCI Economic Conference.. The Integration of Economic Outcome Measures into NCI-Sponsored Therapeutic Trials,* Bethesda MD, Apr. 20-22, 1994.

852. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Eye Institute, *Measuring the Quality of Life for People With Visual Impairment: Proceedings of a Workshop,* M.F. Drummond (cd.), NIH Pub. No. 90-3078 (Bethesda, MD: October 1990).

853. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Eye Institute, *Clinical Trials Supported by the National Eye Institute,* NIH Pub. No. 93-2960, (Bethesda, MD: August 1993).

854. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Hew-t, Lung, and Blood Institute, *The Fifth Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure,* NIH

Pub. No. 93-1088 (Bethesda. MD: January 1 993).

855. U.S. Department of Health and Human Resources, Pub] ic Health Service, National Institutes of Health, National Hew-t, Lung, and Blood Institute, workshop, *National Heart, Lung, and Blood Institute Clinical Guidelines: Planning for the Future*, Rockville, MD, Feb. 22, 1994.

856. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung and Blood Institute, National Asthma Education Program, Expert Panel Report, *Guidelines for the Diagnosis and Management of Asthma*, NIH Pub. No. 91-3042 (Bethesda, MD: August 1991).

857. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, National Cholesterol Educat ion Program, *Second Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II)*, NIH Pub. No. 93-3095 (Bethesda, MD: September 1993).

858. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, National Cholesterol Education Program, Adult Treatment Panel, "Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults," *Archives of Internal Medicine* 148:36-39, *1988.*

*859.* U.S. Department of Health and Human Services, Pub] ic Health Service, National Institutes of Health, National Hew-t, Lung, and Blood Institute, National Cholesterol Education Program, Adult Treatment Panel II, "Summary of the Second Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II)," *Journal of the American Medical Association 269(23):3015-3023,* 1993.

859a. U.S. Department of Health and Human Services, Public Health Scri'ice, National Institutes of Health, National Institute of Diabetes and Digestive and K idney Disease, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 25. 1994.

860. U.S. Department of Health and Human Services, Pub] ic Health Service, National Institutes of Health. National Library of Medicine. *Medical Subject Headings, Annotated Alphabetic List, 1989* (Washington, DC: National Academy Press, 1989).

861. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Office of the Director, *Quality of Life Assessment: Practice, Problems, and Promise: Proceedings of a Workshop, Oct. 15-17, 1990*, NIH Pub. No. 93-3503 (Bethesda, MD: 1990).

862. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Office of the Director, *Guidelines for the Planning and Management of NIH Consensus Development Conferences* (Bethesda, MD: May 1993). '

863. U.S. Department of Health and Human Services. Pub] ic Health Service, National Institutes of Health, Office of Medical Applications of Research, *Health Care Delivery Research Using Hospital Firms, Workshop Summary*, Apr. 30-May 1, 1990 (Bethesda, MD: 1991).

864. U.S. Department of Health and Human Services, Public Health Scr~'ice, National Institutes of Health, Office of Medical Applications of Research, *National Institutes (! f Health Clinical Trials Report FY 1988-FY 1989 (Bethesda,* MD: July 1992).

865. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Office of Medical Applications of Research, workshop on "An Evidencc-Based Health Care System: The Case for Clinical Trials Registries" (comments during genera] audience discussion ), Bethesda, MD, Dec. 6-7. 1993.

866. U.S. Department of Health and Human Ser- L' ices, Public Health Service, National Insti-

(u(es of Health, Office of Medical Applications of Research, Consensus Development Conference, "Past Conferences," unpublished document, Rockville, MD, undated.

867. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, second meeting of the Cost-Effectiveness Panel on Clinical Preventive Services, panel discussion and comments, Washington DC, June 23-24, 1993.

868. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, third meeting of the Cost-Effectiveness Panel on Clinical Preventive Services, panel discussion and comments, Washington DC, Sept. 28-29, 1993.

869. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, Washington, DC, staff personal communication, Nov. 22, 1993.

870. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, fifth meeting of the Cost-Effectiveness Panel on Clinical Preventive Services, panel discussion and comments, Washington DC, Mar. 8-9, 1994.

871. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, Preventive Services Task Force, *Guide to Clinical Preventive Services* (Baltimore, MD: William & Wilkins, 1989).

872. U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, Preventive Services Task Force, "Screening for Adolescent Idiopathic Scoliosis: Review Article," *Journal of the American Medical Association* 269(20):2667-2672, *1993.*

*873.* U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, Preventive Services Task Force, "Home Uterine Activity Monitoring for Preterm Labor: Review Article," *Journal of the American Medical Association* 270(3):371-376, *1993.*

*874.* U.S. Department Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion, Preventive Services Task Force, "Routine Iron Supplementation During Pregnancy: Policy Statement," *Journal of the American Medical Association* 270(23):2846-2854, *1993.*

*875.* U.S. Department of Veterans Affairs, Department of Veterans Affairs Medical Center, Health Services Research and Development Center for Health Care Evaluation, *Department of Veterans Affairs Databases Resource Guide, Version 1.1*, prepared by M.C. Beattie, R.W. Swindle, L.A. Tomko, et al., Palo Alto, CA, May 1993.

875a. U.S. Department of Veterans Affairs, Health Services Research and Development Service, "Cooperative Studies in Health Services: January 1993," working paper, Washington, DC, January 1993.

876. U.S. Department of Veterans Affairs, Veterans Health Administration, Medical Research Service, "CSP Study 298- Treatment of Patients with Acquired Immune Deficiency Syndrome (AIDS) and AIDS-Related Complex," working paper, Washington, DC, undated.

877. U.S. Department of Veterans Affairs, Veterans Health Administration, Medical Research Service, "CSP 995 Trial to Evaluate the Effect of Digitalis on Mortality in Heart Failure," VA-NHLBI, working paper, Washington, DC, undated.

878. U.S. Department of Veterans Affairs, Veterans Health Administration, Medical Research Service, "CSP Study 999a: A Multicenter Clinical Trial of Buprenophine in Treatment of Opiate Dependence," working paper, Washington, DC, undated.

879. "US Finasteride/Doxazosin Trial in BPH," *Scrip World Pharmaceutical News 1898:27, 1993.*

**880,** U.S. General Accounting Office, *Practice* Guide *lines:* The *Experience of Medical Spe - cialty Societies*, PEMD-91 -11 (Washington, DC: U.S. Government Printing Office, 1991 ).

**881,** U.S. General Accounting Office, *Off-Label Drugs: Reimbursement Policies Constrain Physicians in Their Choice of Cancer Thera - pies, GAO;* PEMD-91-*14* (Washington, DC: U.S. Government Printing Office, September 1 991).

**882** U.S. General Accounting Office, *Cross Design Synthesis: A New Strategy for Medical Effectiveness Research,* GAO/PEMD-92-*18* (Washington. DC: U.S. Government Printing Office, 1992).

**883** "The Use of Hormones (Diethylstilboestrol, Estrogen, Ethisterone, Corpus Luteum Hormone ) in the Management of Pregnancy in Diabetics: Report to the Medical Research Council by Their Conference on Diabetes and Pregnancy," *Lancet* 2:833-836, 1955.

**884.** Veatch, R. M., "Consensus of Expertise: The Role of Consensus of Experts in Formulating Public Policy and Estimating Facts," *Journal of Medicine and Philosophy* 16(4):427-445, 1991.

885. Vervoncoeur, C., Research Analyst, Technology Assessment Group, Inc., San Francisco, CA, personal communication, May 25, 1993.

886. Vessey, M. P., Fair-weather, D. V.I., Norman-Smith, B., et al., "A Randomized Double-Blind Controlled Trial of the Value of Stilboestrol Therapy in Pregnancy: Long-Term Follow-Up of Mothers and Their Offspring," *British Journal of Obstetrics and Gynaecology 90: 1007-1017, 1983.*

*887.* Vinicor, F., Cohen, S. J., Mazzuca, S. A., et al., "DIABEDS: A Randomized Trial of the Effects of Physician and/or Patient Education on Diabetes Patient Outcomes," *Journal of Chronic Disease 40: 345-356,1987.*

*888.* Vinokur, A., Bumstein, E., Sechrest, L., et al., "Group Decision Making by Experts: Field Study of Panels Evaluation Medical Technologies." *Journal oj' Personality and Social Psychology 49(1): 70-84, 1985.*

*889.* Wachter, K. W., "Disturbed by Meta-Analysis?" *Science* 241 (4872): 1407-1408, 1988.

890. Wachter, K.W. and Staf, M.L. (eds. ). *The Future of Meta-Analysis (New* York, NY: Russell Sage Foundation, 1990).

891. Waggoner, D. M., Frengley, J. D., Griggs, R. C., et al., "A 'Firm' System for Graduate Training in General Internal Medicine," *Journal of Medical Education 54:556-61,* 1979.

*892.* Ware, J. E., and Sherbourne, C. D., "The MOS 36-Item Short-Form Health Survey (SF-36), I. Conceptual Framework and Item Selection," *Medical Care* 30(6):473-483, *1992.*

*893.* Warlow, C., "Unlocking Datasets," *Lancet 343:* 118, 1994.

894. Warner, K. and Hutton, R., "The Effects of Hospital Cost Containment on the Development and Use of Medical Technology, *Milbank Memorial Fund Quarterly 56:* 187, 1978.

895. Warshafsky, S., Kamer, R. S., and Sivak, S. L., "Effect of Garlic on Total Serum Cholesterol: A Meta-Analysis," *Annals of Internal Medicine* 11 9:599-605, 1993.

896. Wasson, J., Fleming, C., Bruskewitz, R., et al., "The Treatment of Localized Prostate Cancer: What Are We Doing, What Do We Know, and What Should We Be Doing?" *Seminars in Urology 11(1* ):23-26, 1993.

897. Watts, C. A., Jackson, M., and LoGerfo, J. P., "Cost Effectiveness Analysis: Some Problems of Implementation," *Medical Care 17(4):430-434, 1979.*

*898.* Webb, L. Z., Kuykendall, D. H., Zeiger, R. S., et al., The Impact of Status Asthmatics Practice Guidelines on Patient Outcome and Physician Behavior," *Quality Review Bulletin 18:471-476,* 1992.

*899. Webster's Ninth New Collegiate Dictionary* (Springfield, MA: Merriam-Webster, Inc. 1983).

900. Weingart, S. N.. "Acquiring Advanced Technology: Decision-Making Strategies at Twelve Medical Centers," *International Journal of Technology Assessment in Health Care* 9(4):530-538, 1993.

901. Weingarten, S., Director of Health Services, Department of Medicine, Cedar Sinai Hospital, Los Angeles, CA, personal communication, May 18, 1993.

902. Weingarten, S., Agoes, L., Tankel, N., et al., "Reducing Lengths of Stay for Patients Hospitalized with Chest Pain Using Medical Practice Guidelines and Opinion Leaders," *American Journal of Cardiology* 71:259-262, 1993.

903. Weingarten, S., Ermann, B., Bolus, R., et al., "Early Step-Down Transfer of Low-Risk Patients With Chest Pain: A Controlled Intervention Trial," *Annals of Internal Medicine* 113:283-289, 1990.

904. Weingarten, S. Reidinger, M.S., Conner, L., et al., "Reducing Coronary Care Unit Lengths of Stay With a Practice Guideling for Congestive Heart Failure Patients—Insights From A Controlled Clinical Trial," Cedars-Sinai Medical Center, Department of Medicine, Los Angeles, CA., unpublished manuscript, undated.

905. Weingarten, S., Riedinger, M.S., Conner, L., et al., "Practice Guidelines and Reminders to Reduce Duration of Hospital Stay for Patients With Chest Pain Patients: An Interventional Trial," *Annals of Internal Medicine* 120(4): 257-263, 1994.

906. Weinstein, M.D., and Stason, W.B., "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices," *New England Journal of Medicine* 296(13):716-721, 1977.

907. Wells, K.B., Stewart, A., Hays, R.D., et al., "The Functioning and Well-being of Depressed Patients," *Journal of the American Medical Association* 262(7):914-919, 1989.

908. Wennberg, J.E., "On Patient Need, Equity, Supplier-Induced Demand, and the Need to Assess the Outsome of Common Medical Practices," *Medical Care* 23(5):512-520, May 1985.

908a. Wennberg, J.E., School of Medicine, Dartmouth University, Hanover, NH, testimony before the Subcommittee on Health and Environment, Committee on Energy and Commerce, House of Representatives, U.S. Congress, Washington, DC, July 16, 1986.

909. Wennberg, J.E., "Population Illness Rates Do Not Explain Population Hospitalization Rates," *Medical Care* 25(4):354-359, 1987.

910. Wennberg, J.E., "Outcomes Research, Cost Containment, and the Fear of Health Care Rationing," *New England Journal of Medicine* 323(17):1202-1204, July 1991.

***911.*** Wennberg, J. E., "What is Outcomes Research<?", *Medical Innovation at the Crossroads, Vol. I: Modern Methods of Clinical Investigation*, A.C. Gelijns (cd.) (Washington, DC: National Academy Press, 1990).

912. Wennberg, J. E., "Statement on Medical Practice Patterns and Appropriateness of Care," testimony presented before the Committee on Finance, Senate, U.S. Congress, Washington, DC, Oct. 21, 1993.

913. Wennberg, J. E., Director, Center for the Evaluative Clinical Sciences, Dartmouth University, School of Medicine, Hanover, NH, personal communication, April 1994.

914. Wennberg, J. E., and Gittelsohn, A., "Small Area Variations in Health Care Delivery," *Science* 182: 1102-1108, 1973.

915. Wennberg, J. E., and Gittlesohn, A., "Variations in Medical Care Among Small Areas," *Scientific American 246: 120-126,* 129, 132, 134, 1982.

916. Wennberg, J. E., Bunker, J. P., and Barnes, B., "The Need for Assessing the Outcome of Common Medical Practices," *Annual Review of Public Health* 1:277-295, 1980.

917. Wennberg, J. E., Freeman, J. L., and Culp, W.J., "Are Hospital Services Rationed in New Haven or Over-Utilised in Boston?" *Lancet* i: 1185-1189, 1987.

918. Wennberg, J. E., Barry, M. J., and Fowler, F. J., et al., "Outcomes Research, 'PORTS' and Health Care Reform," *Doing More Good than Harm: The Evaluation of Health Care Interventions* K.S. Warren and F. Mosteller (eds.) (New York, NY: New York Academy of the Sciences, 1993).

919. Wennberg, J. E., Blowers, L., Parker, R., et al., "Changes in Tonsillectomy Rates Associated With Feedback and Review," *Pediatrics* 59:821-*826, 1977.*

*920.* Wennberg, J. E., Roos, N. P., Sola, L., et al., "Use of Claims Data Systems to Evaluate Health Care Outcomes: Mortality and Reoperation Following Prostatectomy," *Journal of the American Medical Association* 257:933-936, *1987.*

*921.* Wenneker, M. B., and Epstein, A. M., "Racial Inequalities in the Use of Procedures for Patients with Ischemic Heart Disease in Massa-

chusetts," *Journal of the American Medical Association* 261:253-257, 1989.

922. West, D., Director, Center for Devices and Radiological Health, Food and Drug Administration, Public Health Service, U.S. Department of Health and Human Services, Rockville, MD, personal communication, Mar. 29, 1993.
923. White, C.W., Albanese, M.A., Brown, D.D., et al., "The Effectiveness of Continuing Medical Education in Changing the Behavior of Physicians Caring for Patients With Acute Myocardial Infarction: A Controlled Randomized Trial," *Annals of Internal Medicine* 102:686-692, 1985.
924. White, K.L., "International Comparisons of Health Services Systems," *Milbank Memorial Fund Quarterly* 46:117-125, 1968.
925. White, L.J., Director, Scientific Policy, American College of Physicians, personal communication, Feb. 2, 1994, and Feb. 3, 1994.
926. White, L.J., Director, Scientific Policy, American College of Physicians, letter to the Office of Technology Assessment, U.S. Congress, Washington, DC, Apr. 14, 1994.
927. White, P., Koshy, P., and Duckers, J., "The Management of Pregnancy Complicating Diabetes and of Children of Diabetic Mothers," *Medical Clinics of North America* 37:1481-1496, 1953.
928. Whitehead, A., Whitehead, J., "A General Parametric Approach to the Meta-Analysis of Randomized Clinical Trials," *Statistics in Medicine* 10(11):1665-1677, 1991.
929. Whittle, J., "An Appraisal of the Role of Large Secondary Databases in Medical Technology Assessment," paper prepared under contract for the Office of Technology Assessment, U.S. Congress, Washington, DC, forthcoming 1994.
930. Whittle, J., Section of General Internal Medicine, U.S. Department of Veterans Affairs Medical Center, Pittsburgh, PA, personal communication, Apr. 25, 1994.
931. Whittle, J., Conigliaro, J., Good, C.B., et al., "Racial Differences in the Use of Invasive Cardiovascular Procedures in the Department of Veterans Affairs Medical System," *New England Journal of Medicine* 329:621-627, 1993.
932. Williams, C., "The Value of Consensus" (editorial), *European Journal of Cancer* 27(5):525-526, 1991.
933. Williams, R.L., Chalmers, T.C., Stange, K.C., et al., "Use of Antibiotics in Preventing Recurrent Acute Otitis Media and in Treating Otitis Media with Effusion: A Meta-Analytic Attempt to Resolve the Brouhaha," *Journal of the American Medical Association* 270(11):1344-1351, 1993.
934. Williamson, J.W., Director of Salt Lake REMC, U.S. Department of Veterans Affairs, and Professor of Medicine and Professor of Medical Informatics, University of Utah School of Medicine, Salt Lake City, UT, "Statement," hearing before the Subcommittee on Regulation, Business Opportunities, and Technology, Committee on Small Business, House of Representatives, U.S. Congress, Washington, DC, Oct. 18, 1993.
935. Wilt, T., and Brawer, M., "A Proposal for a VA Cooperative Trial Investigating Prostatectomy versus Observation for Clinically Localized Carcinoma of the Prostate," unpublished manuscript provided by D. Deykin, Cooperative Studies Program, Department of Veterans Affairs Medical Center, Boston, MA, April 1993.
936. Winerip, M., "Trying to Produce Doctors With a Sensitive Touch and a Grasp of Economics," *New York Times*, p. B8, Mar. 2, 1994.
937. Winkler, J.D., Lohr, K.N., and Brook, R.H., "Persuasive Communication and Medical Technology Assessment," *Archives of Internal Medicine* 145:314-317, 1985.
938. Winslow, C.M., Kosecoff, J.B., Chassin, M.R. et al., "The Appropriateness of Performing Coronary Artery Bypass Surgery," *Journal of the American Medical Association* 260(4)505-509, 1988.
939. Winslow, C.M., Solomon, D.H., Chassin, M.R. et al., "The Appropriateness of Carotid Endarterectomy," *New England Journal of Medicine* 318:721-727, 1988.
940. Winslow, R., "Videos, Questionnaires Aim to Expand Role of Patients in Treatment Decisions," *Wall Street Journal*, pp. B1, B3, Mar. 25, 1992.
941. Winslow, R., "Report Card on Quality and Efficiency of HMOs May Provide a Model for Others," *Wall Street Journal*, pp. B1, B6, Mar. 9, 1993.
942. Wolpert, [illegible], "Eliciting and Combining Subjective Judgments about Uncertainty," *Interna-*

*tional Journal of' Technology Assessment in Health Care* 5(4):537-557,1989.

*943.* Woo, B., Woo, B., Cook, E. F., et al., "Screening Procedures in the Asymptomatic Adult: Comparison of Physicians' Recommendations, Patient Desires, Published Guidelines, and Actual Pract ice," *Journal of the Medical Association 254: 1480-1484, 1985.*

*944.* Woolf, S. H., "Practice Guidelines: A New Reality in Medicine I: Recent Developments," *Archives of Internal Medicine 150(9): 1811-* 1818, 1990.

945. Woolf, S. H., "Practice Guidelines: A New Reality in Medicine 11: Methods of Developing Guidelines," *Archives of Internal Medicine* 152:946-952, *1992.*

*946.* Woolf, S. H., "Practice Guidelines: A New Reality in Medicine III: Impact on Patient Care," *Archives of Internal Medicine* 153: 2646-2655, 1993.

947. Woolf, S., Science Advisor, Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Public Health Service, U.S. Department of Health and Human Services, Washington, DC, personal communication, Feb. 16, 1994.

948. Woolf, S. H., and Sex, H. C., "'The Expert Panel on Preventive Services: Continuing the Work of the U.S. Preventive Services Task Force," *American Journal of Preventive Medicine 7(5): 326-330, 1991.*

*949.* Wortman, P. M., Vinokur, A., and Sechrest, L., "Do Consensus Conferences Work? A Process Evaluation of the NIH Consensus Development Program," *Journal of Health Politics, Policy and Law* 13(3):469-498, 1988.

950. Wortman, P. M., and Yeaton, W., "Synthesis of Results in Controlled Trials of Coronary Artery Bypass Graft Surgery" (typescript), Center for Research on the Utilization of Scientific Knowledge, University of Michigan, Ann Arbor, MI, 1982.

951. Yusuf, S., Collins, R., and Pete, R., "Why Do We Need Some Large, Simple, Randomized Trials?" *Statistics in Medicine* 3(4):409-420, *1984.*

*952.* Yusuf, S., "Obtaining Medically Meaningful Answers from an Overview of Randomized Clinical Trials," *Statistics in Medicine* 6(3):281-194, *1987.*

*953. The* Zitter Group and Technology Assessment Group. "1 992 Pharmaceutical Outcomes Activities," San Francisco, CA, 1992.

954. Zook, C. J., and Moore, F. D., '*High-Cost Users of Medical Care," *New England Journal of Medicine* 302:996-1002, *1980.*

# Index


A

ABMT HDC. *See* Autologous bone marrow transplant with high-dosage chemotherapy
Academic Medical Center Consortium
  RAND Corporation cooperative effort, 234
ACC. See American College of Cardiology
Acknowledgments, 203-209
ACP. See American College of Physicians
ACP Observer, 227
Acquired immunodeficiency syndrome. See AIDS
Active approaches to changing physician behavior
  administrative approaches, 180-181, 197
  clinical peer influence, 183
  combining feedback and economic incentives, 186-187, 197
  continuous quality improvement, 187
  economic approaches, 179-180, 197
  feedback, 183-186
  patient influences, 181-183
  practice profiling, 183-186
Activities of Daily Living scales, 40
Acute myocardial infarction
  exclusion of elderly persons in trials, 45
  Physicians' Health Study, 54, 55, 56
  PORT project, 69, 70, 71.72
  thrombolytic drug trials, 51-53, 55, 57, 178
Administrative approaches to changing physician behavior, 180-181, 197
Adolescents. *See also* Infants and children
  preventive services, 221
  scoliosis-related morbidity, 156-157, 158, 224-225
Adrenoleukodystrophy
  Lorenzo's oil treatment, 45
Advisory Committee on Immunization Practices
  clinical practice guidelines, 146, 150, 167, 221-223
Aetna Life and Casualty Co.
  ABMT HDC trial support, 94
  technology assessment activities, 138, 141
African Americans
  coronary bypass surgery data, 47
Agency for Health Care Policy and Research. See *also* Patient Outcomes Research Teams
  activities chart, 13, 136
  BPH trial funding, 84
  budget, 97
  clinical algorithms, 155
  clinical practice guidelines, 2, 9-10, 140-141, 143, 145, 150, 155, 164, 166, 168, 211-214
  comparative effectiveness studies, 82
  components, 96
  cost-effectiveness considerations, 122, 154
  creation of, 2, 23, 34, 73, 96, 135
  evidence rating systems, 156
  Forum for Quality and Effectiveness in Health Care, 140
  guidelines dissemination, 190, 191
  incorporating quality of life measurements into trials, 43
  mandate, 2, 39, 73, 96, 135, 140
  Medical Treatment Effectiveness Program, 33-34,97
  RCT funding, 7
  research coordination efforts, 102-104
  role, 33, 34
  sharing of information with NIH, 84-85
  use of patient representatives on guidelines panels, 155
AHA. See American Heart Association
AHCPR. See Agency for Health Care Policy and Research
AIDS. See *also* HIV
  clinical trials registry, 86
  experimental treatments, 181
  large, simple trials to identify effective treatment, 55
  procedure volume relationship to patient outcome, ***29***
  ***SF-36*** quality-of-life scale use in trials, 83
  surrogate endpoint use in drug trials, 91
  zidovudine treatment study, 83
ALD. *See* Adrenoleukodystrophy

ALLHAT clinical trial of hypertension and cholesterol-lowering treatments, 85, 87
AMA. *See* American Medical Association
Ambulatory Sentinel Practice Network, 88-89
American Academy of Family Practice, 88
American Academy of Pediatrics
  immunization recommendations, 169
  infectious disease control monograph, 225-226
  statement on appropriate use of tetracycline, 30
American Cancer Society
  colorectal cancer screening, 166
American College of Cardiology
  clinical practice guidelines, 192
American College of Physicians
  Clinical Efficacy Assessment Project, 135, 146, 150, 151, 164,226-228
  clinical practice guideline process, 226-228
  colorectal cancer screening, 166
  Medical Necessity Program cooperation, 139
  technology assessment activities. 141
American College of Radiology
  Medical Necessity Program cooperation, 139
American College of Surgeons
  Medical Necessity Program cooperation, 139
American Diabetes Association
  collaboration with CDC's ACIP, 221
American Gastroenterological Association
  colorectal cancer screening, 166
American Heart Association
  clinical practice guidelines, 192
American Hospital Association
  new technologies commentary periodical, 138
American Lung Association
  collaboration with CDC's ACIP, 221
American Medical Association
  clinical practice guidelines, 36, 229-231
  Diagnostic and Therapeutic Technology Assessment Program, 136-137, 146-147, 150, 151, 153,229-231
  *Directory of Practice Parameters*, 139
  opposition to NCHCT standards, 134
  *Practice Parameter Update, 229*
  technology assessment activities, 141
  Teflon® treatment for urinary incontinence, 170
American Psychological Association
  and AHCPR'S depression guidelines, 152
American Society for Gastrointestinal Endoscopy
  colorectal cancer screening, 166
American Society of Anesthesiologists
  anesthesia monitoring guidelines, 193-194
American Society of Hospital Pharmacists
  study of controlled formularies, 126
American Urological Association
  BPH trial funding, 84
AM I. *See* Acute myocardial infarction
Anesthesia monitoring guidelines, 193-194
Anisoylated plasm inogen-activator complex
  acute myocardial infarction treatment, 52-53
*Annals of Internal Medicine*, 227,228
Appropriateness of care
  "appropriate" definition, 30-31
  cost of care and, 5
  RAND study, 30-33, 154, 168
APSAC. *See* Anisoylated plasminogen-activator complex
Arthritis
  knee replacement surgery for elderly persons, 71
  opinion leaders influence on treatment, 183
  patient education, 181
Aspirin
  role in preventing heart attacks, 54, 55, 56
ASPN. *See* Ambulatory Sentinel Practice Network
Asthma
  clinical practice guidelines, 146, 217, 218
Australia
  CEA standardization efforts, 128
Autologous bone marrow transplant with high-dosage chemotherapy, 93,94, 181
AZT. *See* zidovudine

B

Back injuries
  hospitalization for, 27
Back pain
  PORT project, 69, 70,72
  surgical treatment, 184-185
  video disc for patient education, 182
Bakketeig
  firms trial study, 59
Basic tools for measuring health status and outcomes
  attributes, 40
  mortality, 40
  patient self-assessment and, 40-41
  quality of life measurements, 41-42
Bayesian approach to meta-analysis, 64
Benign prostatic hyperplasia. See *also* Prostate disease
  clinical trials funding, 83
  comparative outcomes of procedures, 51
  mortality rates, 49
  patient education efforts, 181-182
  patient self-assessment, 40
  PORT project, 84-85, 157
  treatment study, 28-29
Beta-Blocker Heart Attack Trial
  per-patient cost, 54
Beta-blocking drugs
  use after heart attack, 50, 54

Beta-carotene
cancer prevention role, 54, 55, 56
Blue Cross and Blue Shield Association
ABMT/HDC trial support, 94
clinical practice guidelines in Illinois, 36
cooperative effort with Kaiser Permanence Medical Care Program, 138, 143
Medical Necessity Program, 138, 139
Technology Evaluation and Coverage program, 138, 139
Bone marrow transplant. See Autologous bone marrow transplant with high-dosage chemotherapy
BPH. *See* Benign prostatic hyperplasia
Breast cancer
ABMT/HDC treatment, 93,94, 181
age-specific survival rate study, 64
Brown and Fintor CEA for early detection, 112-113
mammography screening CEA, 123
meta-analysis of alcohol in relation to, 62
Brown and Fintor
CEA for early detection of breast cancer, 112-113, 115
B yar
database study criticism, 46

C

Cal ifornia
use of clinical practice guidelines, 36
Canada
CEA standardization efforts, 128-129
Cesarean delivery rate compared with United States, 193
clinical practice guidelines, 36
Cardiac angiography
osmolality of radiographic material, 117
Cardiac pacemakers
decreased use of, 192-193
Cardiac surgery
mortality data, 186
Caret id endarterectomy
appropriateness of, 30, 32
mortality following, 47, 165
Cataracts
AHCPR'S clinical practice guidelines, 152,213
PORT project, 68,69,71, 72,74, 82
retinal detachment and, 48-49, 71, 72, 74, 82
CCOP. See Community Clinical Oncology Program
CDC. *See* Centers for Disease Control and Prevent ion
CEA. See Cost-effectiveness analysis
CEAP. *See* Clinical Efficacy Assessment Project
Centers for Disease Control and Prevention
Advisory Committee on Immunization Practices, 146, 150, 167,221-223
components, 100
cost-effectiveness analyses, 123-124
how-to manual on decision analysis, cost measurement, and CEA, 123
mission, 13, 99
*Morbidity and Mortality Weekly Report, 166-167, 221*
prevention effectiveness focus, 99-100
vacc ination recommendations, 140
Cesarean del ivery. See *also* Pregnancy
c1 inical practice guidelines, 164
current practice profile, 185
rate comparison in United States and Canada, 193
rates in managed care settings, 179
Chalmers and colleagues
meta-analysis reliability study, 61-62, 63
CHAMPUS insurance program, 135
Chassin
practice variation, 37
Children. See Infants and children
Chlamydia trachomatis guidelines, 221
Cholesterol reduction
ALLHAT study, 87
clinical practice guidelines, 146, 217, 218
meta-analysis study, 62
CIGNA Corp.
ABMT/HDC trial support, 94
technology assessment division, 138
Cleveland Health Qua] ity Choice Project
report card program, 24
Cleveland Metropolitan General Hospital
firms trial example, 57
Clinical algorithms, 147, 148, 150, 155, 161, 232-233
Clinical-economic trials
"blinding" of patients and physicians, 116-117
clinical protocol influence on resource use, 116-117
cost measurement difficulties, 118
cost variability and sample size, 117-118
examples, 117
generalizability, 118
goals, 116
Clinical Efficacy Assessment Project
ACP project for internal medicine, 135
clinical practice guidelines, 146, 150, 151, 226-228
evidence basis for guidelines, 164
Clinical management guidelines, 195

Clinical practice guidelines. *See* also appendix C; Physician behavior; *specific federal government agencies and private organizations by name*
'basis for recommendations, 164-166
clinical algorithms, 147, 148, 150, 155, 161, 232-233
core tasks descriptions, 210-236
cost and cost-effectiveness issues, 153-154
defensive medicine and, 35
definitions, 24-26, 139-140
development-related issues overview, 147, 149
federal efforts, 9-11, 145-146
group interaction and decisionmaking techniques, 11, 157-164
guideline implementation, 190-197
guideline panel selection, 151-153
identifying and synthesizing evidence, 155-157
impact of guidelines on practice, 173-198
link between methods and recommendations, 168-169
malpractice insurance and, 35
methodological research, 11, 15.168
mode] for basing health insurance benefits on, 36
organization and administration of guideline activities, 166-168
overlap of recommendations, 168-169
patient health status and functioning, 154-155
patient preferences, 155
policy options, 15-16
prioritizing guideline topics, 169-170
private efforts, 146-147
research needs, 170-171
scope, 153-155
state legislative activities, 35
support for, 34
technology assessment and, 9-10, 139-141
topic selection criteria, 149-151
Clinton administration
effectiveness and cost-effectiveness research and, 36
health care reform proposal, 34
insurance coverage of new and experimental technologies, 92
Cochrane, Archie
randomized controlled trials, 19-20
Cochrane Collaboration
centers, 81
database development, 171
features, 79-80
Codman, Earnest A.
patient outcomes research, 19
Colonoscopy
rates of, 186
Colorectal cancer
screening for, 166-167
Combinability of results issue in meta-analysis, 62-63
Community-based research
administrative barriers, 88
examples, 87-90
Community Clinical Oncology Program, 88
Comparability of cost-effectiveness analyses, 112-114
Comparative effectiveness trials, 82-83
Comparative uses of administrative databases, 48-51
Congestive heart failure
digitalis treatment, 55,82-83
hospital length-of-stay guidelines, 194-195
Connecticut
geographic variation in inpatient care utilization, 27
hospitalization rates for back injury, 27
Consensus Development Conference Program of NIH
basis for guidelines, 164
clinical practice guidelines process, 214-217
direct mailings, 190
establishment of, 133
hearing impairment screening for infants, 153
methodology of research, 168, 170
physician practice change, 190, 192
public forums, 155
purpose, 141, 146
Continuous quality improvement, 187, 198
Coronary angiography
geographic variation in use, 32
Coronary artery disease
treatment comparison, 49-50
Coronary bypass surgery
rates for African Americans, 47
Cost-benefit analysis
description, 108
Cost-effectiveness analysis
analytic approaches, 107-108
clinical-economic trials, 116-118
comparability issue, 8, 112-114
conclusions, 127-130
cost-utility analysis, 118-122
definition, 23-24
description, 7, 108
federal activities, 9, 13, 122-125
international standardization efforts, 128-129
mechanics of, 11O-111
models for, 111
origins, 109
policy options, 14
private sector activities, 8, 125-127
purposes, 108-109
quality issue, 114-115
technology assessment comparison, 132

uses of, 108-110
Cost-Effectiveness Panel on Clinical Preventive Services
ODPHP sponsorship of, 101
Cost-utility analysis
advantages, 118-119, 127
dangers of, 120-121.127-128
description, 8, 108
distributional considerations, 120-121
measuring utilities, 119-120
social decisions and, 121
Council on Health Care Technology Assessment
establishment, 134-135
CQI. See Continuous quality improvement
CRC. See Colorectal cancer
CUA. *See* Cost-utility analysis

**D**

Daniels and Schroeder
patient outcomes research, 28
Database studies
comparative uses, 48-51
contributions, 74-75
descriptive uses, 46-48
RCT comparison, 5, 46
DATTA. See Diagnostic and Therapeutic Technology Assessment Program
Davis and colleagues
strategies to change physician behavior, 189-190
Decision analysis
advantages and disadvantages, 65-66, 75
description, 65, 162-163
Decision support systems, 159-162
Decisionmaking
factors, 175-177, 178
group interaction and, 159-162
models, 173-175
rules and procedures, 162-163
Defensive medicine
clinical practice guidelines and, 35
Delphi technique, 159, 160-161, 171,233,235-236
Department of Defense
CHAMPUS insurance program, 135
research role, 21
Department of Health and Human Services. See *specific agencies by name*
Department of Veterans Affairs. See also Veterans Health Administration
BPH trial funding, 84-85
clinical trials registry, 86
comparative effectiveness studies, 82
Health Services Research & Development Service, 102
Medical Research Service, 82,88, 102
Prostate Cancer Intervention Versus Observation Trial, 82
Rehabilitation Research and Development Service, 102
role, 13, 21
Depression
American Psychological Association's disagreement with AHCPR'S guideline, 152
patient outcome effect, 30
DES. *See* Diethylstilbestrol
Descriptive uses of administrative databases, 46-48
Detailing form of peer influence, 183
Diabetes
insulin therapy trial, 46
laser treatments for diabetic retinopathy, 178-179
patient involvement in treatment, 181
Diagnostic and Therapeutic Technology Assessment Program
clinical practice guidelines, 146-147, 150, 151, 153,229-231
establishment of, 136-137
Dickersin and Berlin
meta-analysis studies, 62
Diethylstilbestrol
inadequate research example, 20, 78
Digitalis
congestive heart failure treatment, 55, 82-83
per-patient cost of study, 54
Dingell, Rep. John, 3, 199
*Directory of Practice Parameters,* 139
Drug prescribing guidelines, 195. See *also* Food and Drug Administration; Pharmaceutical companies
Duke Database for Cardiovascular Disease, 49-50

**E**

Ear infections
acute ear infection management trial, 103
Economic approaches to changing physician behavior, 179-180, 197
ECR1
clearinghouse activities, 138
Eddy
clinical practice guideline definition, 139-140
Effectiveness research. See also Cost-effectiveness analysis
appropriateness of care, 5, 30-33
clinical research infrastructure improvement, 87-90
comparative evaluation of new technologies, 90-95
comparative trials, 82-83
conclusions, 37-38, 104-105
definition, 23
federal funding and support, 4-5, 13,95-104

federal medical treatment effectiveness program, 33-34
gaps in existing federal effort, 77-95
geographic variations in medical practice. 26-28
health care reform and, 34-37
history, 26
ongoing trials registry, 85-87
patient outcomes, 28-30
quality-of-life assessment inclusion, 83-85
research coordination, 102-104
systematic reviews, 78-82
Efficacy and effectiveness comparison, 23
Elderly persons
arthritis and, 71
CEA of community-based care. 122
exclusion of in AM I trials, 45
knee replacement surgery PORT, 71
Medicare claims to estimate costs of health care, 46
Pap screens study, 111
Eli Lilly and Co.
cost-effectiveness analyses, 126
Emphysema
surgical treatment, 22
European Union
CEA standardization efforts, 129
EuroQol index, 42
Evaluation of efficacy and safety
DES example, 20
federal government role, 20-21
reasons for pub] ic lack of understanding, 21
surgical technique examples, 22
terminology, 23-26
Experimental studies
description, 43-46
Experimental technologies. *See* Insurance coverage of new and experimental technologies

F

FDA. *See* Food and Drug Administration
Fecal occult blood testing
colorectal cancer screening, 166
Federal government. *See* also *specific agencies by name*
cost-effect iveness activities, 122-125
directing and coordinating activities, 16-17
effectiveness research, 33-34, 77-105
guidel ine activities, 210-225
health technology assessment role, 2-3, 13-17, 131-135
Feedback influence on physician behavior, 183-187, 198
Femur fracture
procedure volume relationship to patient outcome, 29
Firms trials
advantages. 57
applications, 7, 58-59
basic concepts, 57
community-based research possibility, 89-90
limitations, 57-58
methodological issues, 58
Food and Drug Administration
medical dev ice evaluation, 91
meta-analysis uses. 64
new technology evaluation role, 91-92
rev iew of cost-related claims in drug advertising, 9, 127
role, 21
surrogate endpoints in drug trials, 91
transition of drugs from prescription to over-the-counter, 92
Formularies, 126, 180
Forum for Quality and Effectiveness in Health Care, 140
Foundation for Informed Medical Decisionmaking
v ideo disc technology for patient education, 181-182

G

Gallstone lithotripsy
CEA of. 122
General Accounting Office
meta-analysis uses, 64
Geographic variation in medical practice
appropriateness ratings and, 31-33
high-cost procedure use, 179-180
surgical procedure studies, 26-28
GISSI. *See* The Gruppo Ital iano per lo Studio della Streptochinasi Nell' Infarto Miocardio
Glaxo, Inc.
cost-effectiveness analyses, 126
Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries. 52-53
Glover
tonsillectomy study, 26
GM-CSF. *See* Granulocyte macrophage colony stimulating factor
Granulocyte macrophage colony stimulating factor
Hodgkins disease adjuvant therapy, 117
Grassley, Sen. Charles E., 3, 199
Group interaction and decisionmaking
decision support systems, 159-162
formal group processes, 159

The Gruppo Italiano per 10 Studio della Streptochinasi Nell 'Infarto Miocardio, 52-53, 75, 87
*Guide to Clinical Preventive Services*, 146, *224,225*
GUSTO. *See* Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries

H
Hadorn
model for basing health insurance benefits on, 36
Harvard Community Health Plan
clinical algorithms, 147. 148, 150, 155, 232-233
clinical practice guidelines process, 231-234
cost-effectiveness considerations, 154
CQI methods to change practice, 187, 188
methodology of research, 168
Hatch, Sen. Orrin, 3,199
HCFA. See Health Care Financing Administration
HCHP. *See* Harvard Community Health Plan
Health Care Financing Administration
administration of Medicaid and Medicare programs. 101, 124
basis for guidelines, 164
cost-effectiveness analyses, 13, 124-125
database studies, 50-51
linkage of SEER-Medicare databases, 101
Office of Research and Demonstrations, 101
purpose, 13, 124
Health care reform
effectiveness research and, 34-37
Health effects evaluation techniques
basic tools, 40-43
primary studies, 43-59
secondary techniques. 59-67
Health Industry Manufacturer's Association
opposition to NCHCT standards, 134
Health Insurance Association of America
payment for new and experimental treatments, 93
technology assessment activities, 138-139
Health maintenance organizations. *See also specific organizations by name*
collaborative effort on prevention strategies, 100
controlled formularies, 126
*Healthy People 2000* report, 101
Health technology assessment
definitions, 9, 131-132
federal efforts, 9-10, 15, 131-135
organizations involved in health technology assessment activities, 137
origins of, 133
private sector efforts, 9, 135-139
transition from public to private sector, 141-142
Health Utilities Index, 42
Hearing impairment screening for infants, 153
Heart attack. *See* Acute myocardial infarction
Henry and Wilson
meta-analysis reliability study, 61
Hepatitis B vaccines
cost-effectiveness study, 115
Hershey and colleagues
firms trials study, 58
High blood pressure. *See* Hypertension
Hip fracture
PORT project, 72
Hip replacement
clinical-economic trial of hospitalization after, 117
HIV
AHCPR guideline, 212, 213, 214
HMO Group
TEMINEX technology>' assessment project, 138
HMOs. *See* Health maintenance organizations
Hodgkins disease
clinical-economic trial on benefits of GM-CSF, 117
Hospital Association of New York
technology assessment manual for hospitals, 138
Hospitalization
for back injury, 27
geographic variation in inpatient care utilization, 27
length-of-stay guidelines, 194-195
Medicaid prescription drug reimbursement limits. 180-181
for pneumonia, 71-72
relationship to physician reimbursement, 179
Human immunodeficiency virus. *See* HIV
Hypertension
clinical practice guidelines, 146, 217, 218
comparative effects of treatments study, 83, 85
Hysterectomy rates
efforts to reduce, 183, 184-186
geographical variation. 26-27

I
ICUs. *See* Intensive care units
Illinois
Blue Cross and Blue Shield clinical practice guidelines, 36
Mt. Sinai Hospital Cesarean delivery rates. 193
Immunizations. See *also* Vaccinations
American Academy of Pediatrics recommendations, 169
CDC recommendations, 146.150, 167.221-223
Impact of guidelines on practice
active approaches to changing behavior, 179-187
factors influencing. 195-197
guideline implementation, 190-197

literature on physician behavior change, 187, 189-190, 198
passive approaches to changing behavior. 177-179, 198
physician behavior, 173-177
Indiana
Reganstrief Health Center firms trials study, 58-59
Infants and children. See also Adolescents; Cesarean delivery: Pregnancy
acute ear infection management trial,103
community-based research, 89
hearing impairment screening, 153
immunization recommendations.169
Informed consent, 181
Institute of Medicine
clinical practice guidelines, 34, 139, 140, I 41, 142, 150
cost-effectiveness considerations.154
Council on Health Care Technology Assessment, 134-135
database studies, 47
*Medical Technology Assessment Dirt' ctory*,136, *138*
Insurance coverage of new and experimental technologies. See *also specific insurance companies by name*
autologous bone marrow transplant with high-dosage chemotherapy example. 93,94
concerns, 94, 95
"medically necessary" definition, 92-93
Intensive care units
length-of-stay guidelines, 194-195
Intermountain Health Systems
CQI methods to change practice, 187, 188
International Study of Infarct Survival, 51-53, 87
IOM. See Institute of Medicine
ISIS. See International Study of Infarct Survival
Italy
acute myocardial infarction study, 52-53, 75, 87

J
JCAHO. See Joint Commission on Accreditation of Healthcare Organizations
Jefferson and Demichelli
literature review on cost-e ffectiveness of *hepatitis B* vaccines, 115
Joint Commission on Accreditation of Healthcare Organizations
Indicator Monitoring System, 25

K
Kaiser Permanence Medical Care Program
cooperative effort with Blue Cross and Blue Shield Association, 138, 143
report card project, 25
Karnofsky Index. 40
Kennedy, Sen. Edward, 3, 199
Klawansky and colleagues
meta-analysis study, 64
Knee replacement surgery
PORT project, 71

L
Large, simple trials
characteristics, 51
current applications, 7, 55, 57
GISSI example, 52-53,75, 87
head-to-head comparisons of treatments, 53
implications and limitations, 53-55, 75
ISIS example, 51-53, 87
need for simplicity in design, 54-55
per-patient cost advantage, 54
Laser capsulotomy. See Cataracts
Lau, Antman, and colleagues
AM I meta-analysis, 78
Law and colleagues
meta-analysis study of reducing cholesterol, 62
Leonard Davis Institute
CEA guidelines development, 127
Lidocaine
in myocardial infarction treatment, 49, 50, 78-79
"Linking Medical Evidence with Clinical practice" workshop, 201, 204
Lomas and colleagues
Cesarean delivery rate comparison, 193
Longitudinal Comparisons of Measures for Health Outcomes, 122
Lorenzo's oil, 45
Luft and colleagues
procedure volume relationship to patient outcome, 29
Lusted
decision analysis paper, 65

M
Maine
back surgery PORT, 70
cl inical practice guidelines legislation, 35
hospitalization rates for back injury, 27
surgical procedure reduction study, 184-185

Maine Medical Assessment Foundation, 184-185
Malpractice insurance
clinical practice guidelines and, 35
Mammography
CEA of breast cancer screening, 123
Markov model
use in cost-effectiveness analysis, 111
Maryland
clinical practice guidelines legislation, 35
Cochrane Collaboration center, 81
insurance coverage of new and experimental technologies. 92
Maryland Quality Indicator Project
report card program, 24
Mason and colleagues
comparative study of cost per QALY of medical technologies, I I3-114
Massachusetts
anesthesia monitoring guidelines, 193-194
geographic variation in inpatient care utilization, 27
Matchar
benefits of decision analysis, 65
Medical devices
FDA role in evaluating, 91.132-133
Medical Necessity Program, 138, 139
Medical Outcomes Study, 29-30, 41, 42
Medical school curricula
decisionmaking training, 190
*Medical Technology Assessment Directory*, 136, 138
Medical Treatment Effectiveness Program, 33-34, 97
Medicare. See also Health Care Financing Administration
database studies, 51
diagnosis-based payment for inpatient services, 179
effort to link with SEER registry, 99
use of claims data to estimate costs of health care for elderly persons, 46
utilization reviews, 190
MEDLINE®
AIDS treatment trials, 86
augmented clinical trials database for, 82
constraints, 81
MEDTEP. *See* Medical Treatment Effectiveness Program
Meta-analysis. *See also* Systematic reviews
Bayesian approach, 64
characteristics, 59-61
combinability of results issue, 62-63
cross-design syntheses, 64
fixed effects model, 63
nine steps of, 60-61
pooled analyses comparison, 61
protocol issue. 63-64
publication bias issue, 63
random effects model, 63
topics for, 64
MetaWorks, Inc.
activities. 138
Metropolitan Life Insurance Co.
ABMT/HDC trial support, 94
Minnesota
clinical practice guidelines legislation. 35
technology assessment consensus document, 138
MMAF. *See* Maine Medical Assessment Foundation
*Morbidity and Mortality Weekly Report*, 166-167, 221
MOS. *See* Medical Outcomes Study
Mt. Sinai Hospital (Chicago) Cesarean delivery rates, 193
Multiple sclerosis
newly approved drug for, 182

**N**

National Association of Insurance Commissioners
working group on payment for new and experimental treatments. 93
National Asthma and Prevention Education Program, 217, 218
National Cancer Institute
ABMT HDC trial, 93, 94
clinical-economics study funding, 117
clinical practice guidelines, 146, 150, 164, 167, 219-221
colorectal cancer screening, 166
community-based research. 88
cost-effectiveness considerations, 123, 154
database studies, 50
PDQ database, 86, 146, 154, 167-168
screening recommendations, 140
SEER registry, 99
supportive cancer care, 154
National Center for Health Care Technologies." I 33-134
National Center for Health Services Research
clinical-economic studies, 117
Office of Health Technology Assessment, 134
outcomes research program, 33
National Center for Health Statistics, 100"
National Center for Nursing Research, 99
National Cholesterol Education Program. 217, 218
National Committee for Quality Assurance
report card project. 25
National Eye Institute
ongoing studies, 99
quality-of-life measurements, 83, 84

National Foundation for Infantile Paralysis
polio vaccine study, 51
National Health Information Center, 101
National Heart, Lung, and Blood Institute
basis for guidelines, 164
cholesterol, asthma, and blood pressure guidelines, 140
clinical practice guidelines, 146, 150, 151, 167, 217-219
cost-effectiveness analyses, 123
digitalis treatment study, 55,82-83
literature reviews, 155
patient representatives, 155
National High Blood Pressure Education Program, 217, 218
National Institute for Allergies and Infectious Disease
incorporation of quality -of-] ife measurements in trials, 83
National Institute for Child Health and Human Development
acute ear infection management trial, 103
meta-analysis study, 79
National Institute of Mental Health
CEA studies, 123
co-sponsorship of schizophrenia PORT, 103
National Institute on Aging
cost-effectiveness analyses, 123
National Institutes of Health. See also Consensus Development Conference Program; Office of Medical Applications of Research; *specific institutes by name*
activities, 13
BPH trial funding, 84-85
budget, 97,98
clinical trials registry, 86-87
comparative effectiveness studies, 82
components, 98
cost-effectiveness activities, 122-123
database development, 171
effort to link Medicare with SEER registry, 99
features of studies. 97,99
incorporating qual ity of life measurements into trials, 43
role, 20-21
sharing of information with AHCPR, 84-85
trialsponsorship, 79, 82
use of quality-of-life measurements, 83-84
National Library of Medicine, 81, 171,212
NCHCT. See National Center for Health Care Technologies
NCHSR. See National Center for Health Services Research
NEI. See National Eye Institute
New technology evaluation. See also Insurance coverage of new and experimental technologies
avenues for enrollment in comparative evaluations, 90
enhancing number and quality of, 90-91
FDA role, 91-92
New York
cardiac surgery outcomes data, 24
Health Department's cardiac bypass surgery data form, 186
NHLBI. See National Heart, Lung, and Blood Institute
NICHD. See National Institute for Child Health and Human Development
NIH. See National Institutes of Health
NIMH. *See* National Institute of Mental Health
NLM. See National Library of Medicine
Nominal Group Technique, 159
Norway
health care priorities study, 121
Nottingham Health Profile, 42

0

Observational studies
description, 43
Odone, Augusto and Michaela, 45
Office of Disease Prevention and Health Promotion. *See also* U.S. Preventive Services Task Force
activities, 13, 100-101
budget, 100
cost-effectiveness analyses, 125
*Healthy People 2000* report, 101
Office of Health Technology Assessment
health technology assessment activities, 133
policy options, 15
proposal to augment funds with private funds, 135
purpose, 134, 143
Office of Malaria Control in War Areas, 99
Office of Medical Applications of Research, 86, See also Consensus Development Conference Program
clinical guideline process, 146, 150
establishment of, 133
Office of Technology Assessment
health technology assessment activities, 133
purpose of, 133
study profile, 3, 199-202
technology assessment definition, 131-132
Ohio
Cleveland Metropolitan General Hospital firms trial example, 58
Wade Park Veterans Affairs Medical Center firms trial example, 58

OHTA. *See* Office of Health Technology Assessment
OMAR. *See* Office of Medical Applications of Research
Ongoing clinical trials registry
barriers to, 86-87
purposes, 85-86
Opinion leader\' influence on physician behavior, 183
Oregon
prioritization of health care services, 35, 42.120. 130
Outcomes-based management. *See also* Patient outcomes"
definition, 23
report cards, I 2, 24-25
Outcomes research. *See also* Patient outcome\"
definition. 23

**P**

Passive dissemination of information, 177-179, 198
Patient education, 181-1 83, 197
Patient involvement in decision making, 181-183
Patient outcomes
Medical Outcomes Study, 29-30
physician preferences and uncertainty as factors, 28
procedure volume relationship to outcome, 29
prostate disease research, 28-29
report card programs, 24-2
Patient Outcomes Research Teams
AHCPR funding, 73, 79
components, 67-68
conference conclusion s." 73
contributions. 6, 68, 70-72.75-76
cost-effectiveness analysis and. 122
description, 66-67
future plans, 73
goals, 67
interPORT work groups, 68
limitations and frustrations, 72-73
planned and current projects, 69
structure, 67
topics addressed, 67
Patient reminders, 1 90, 192
PDQ database. *See* Physician Data Query database
Peer influence on physician behavior, 183
Peripheral artery angioplasty, 47
Peritonitis
antibiotic treatment of spontaneous bacterial peritonitis, 117
Pharmaceutical companies
cost-effectiveness analy ses, 125-1 26
marketing directly to patients, 181
Pharmacoeconomics, 111-112, 125
Physician behavior
active approaches to changing, 1 1, 1 79- 1 87, 1 98
behavioral models, 173-175, 176
conclusions, 197-198
factors underling current practices. 1I, 175, 177, 178
impact of guidelines on, 1 90, 192-197, 198
literature on, 187, 189-190
opinion leaders influence. 11. 183
passive approaches to changing. 11-12, 177-179, 198
Physician Data Query database. 86, 146, 154, 167-168, 220-221
Physician ownership of facilities. 179
Physician Payment Review Commission
clinical practice guidelines and, 34
Physicians' Health Study
description, 56
large, simple trial example, 55
per-patient cost, 54
PIVOT. *See* Prostate Cancer Inter\' ention Versus Observation Trial
Pneumonia
PORT project, 71-72
Polio vaccine, 5 I
Pooled analysis
meta-analysis comparison, 6 I
PORTs. *See* Patient Outcomes Research Teams
*Practice Parameter Update*, 229
Practice profiling influence on physician behavior, 12, 183-186, 198
Pregnancy. *See also* Cesarean delivery
Cochrane Collaboration on premature births, 80
large, simple trials for fetal movement counting, 55
Preventive care guidelines, 192
Primary studies to evaluate health effects
categories, 43-46
database studies, 46-51
firms trials, 57-59
large, simple trials. 5 I -57
Private sector. *See also specific organizations by name*
cost-effectiveness analysis, 125-127
guideline activities, 225-236
health technology assessment, 9, 135-139
Procedure volume relationship to patient outcome, 29
Prospectiv e Pay ment Assessment Commission, 134
Prostate Cancer Intervention Versus Observation Trial. 82

Prostate disease. *See also* Benign prostatic hyperplasia
CEA of prostate cancer screening, 123
PORT project, 69,70,72, 84-85
screening for, 181
Prostatectomy. See Benign prostatic hyperplasia; Surgery; Transurethral prostatectomy
Prudential Insurance Co. of America
ABMT/HDC trial support,94
technology assessment division, 138
Publication bias issue in meta-analysis,63

*Q*
QALYs. See Quality-adjusted life years
Quality-adjusted life years
inconsistencies in measurement and use of, 113, 115-116
value of life, 121
Quality of cost-effectiveness analyses, 114-115
Quality-of-life measurements
applications and limitations, 42-43
incorporating in clinical trials, 83-85
Quality of Well Being Scale, 41,42, 119
Quinidine
FDA trial, 91-92
QWB Scale. *See* Quality of Well Being Scale

R
RAND Corp.
Academic Medical Center Consortium cooperative effort, 234
appropriateness of care, 30-33, 154, 168
basis for guidelines, 164
clinical practice guidelines, 141, 147, 153, 234-236
cost-effectiveness considerations, 154
decisionmaking support, 160-161, 163
Delphi group process method, 171,233,235-236
Health Insurance Experiment, 179
ratings of appropriateness, 31
Randomized controlled trials
AHCPR funding, 7
criticisms of, 44-46
database study comparison, 46
description, 43-44
impact on physician behavior, 178
internal validity, 44
underuse in government's effectiveness initiative, 6-7,74-75
Ray and colleagues
tetracycline appropriateness study, 30
RCTS. See Randomized controlled trials
Reganstrief Health Center
firms trials study, 58-59
Report cards for patient outcomes, 12,24-25
Rhode Island
surgery rates, 27
Risk Management Foundation of Harvard Medical Institutions
anesthesia monitoring guidelines, 193-194
Roper, William
effectiveness research proposal, 33
Rubin
meta-analysis study characteristics, 63-64

S
Salivary glands
surgery to protect glands during radiation therapy, 22
Salk polio vaccine study, 51
Schizophrenia
PORT project, 103
Scoliosis
morbidity among adolescents, 156-157, 158, 224-225
Secondary techniques to synthesize research results
decision analysis, 65-67
meta-analysis and other systematic reviews, 59-65
SEER registry. See Surveillance, Epidemiology and End Results registry
Sexually transmitted diseases. See *also* AIDS; HIV
clinical practice guidelines, 221
SF-36 patient functioning questionnaire, 30,41,42, 83
Sickle cell disease
AHCPR guideline,212
Sickness Impact Profile, 41, 42,83
SIP. See Sickness Impact Profile
Soumerai and colleagues
feedback and reminder systems study, 190
Spain
CEA standardization efforts, 129
Streptokinase
acute myocardial infarction treatment, 52-53, 55, 57,78
Stroke
PORT project, 71, 120
*Surgeon General's Report on Nutrition and Health,* 101
Surgery. See *also specific types of surgery by name*
anesthesia monitoring guidelines, 192, 193-194
benign prostatic hyperplasia treatment, 28-29
cardiac surgery mortality data, 186
emphysema treatment, 22
hysterectomy rates, 26-27
to preserve salivary glands during radiation therapy, 22

procedure volume relationship to patient outcome, 29
tonsillectomy studies, 26, 27
wound infection study, 188
Surveillance, Epidemiology' and End Results registry, 99
Systematic reviews. *See* also Meta-analysis
Cochrane Collaboration, 79-81
contributions, 78-79
improving efficiency of, 13, 81-82

**T**

Technology Assessment Group, Inc., 138
Technology Evaluation and Coverage program, 138, 139
*Technology Reports,* 138
Teflon ®
urinary incontinence treatment, 170
TEMINEX technology assessment project, 138
Tennessee
tetracycline appropriateness study, 30
Terms and definitions, 23-26
Test ordering guidelines, 195
Tetracycline
appropriateness of use study, 30
Thrombolytic dregs
effects on acute myocardial infarction, 51-53, 55, 57, 178
Tissue-type plasminogen activator
acute myocardial infarction treatment, 52-53, 55, 57
Tonsillectomy studies, 26,27,49
TPA. *See* Tissue-type plasm inogen activator
Transfusion practice guidelines, 195
Transurethral prostatectomy
benign prostatic hyperplasia treatment, 28-29,84, 157
mortality rate, 47, 49
Travelers Corp.
ABMT/HDC trial support, 94
Tufts University School of Dental Medicine
surgery to protect salivary glands during radiation treatment, 22
TURP. See Transurethral prostatectomy

**u**

Udvarhelyi and colleagues
fundamental principles of cost effectiveness, 114-115
United HealthCare Corporation
report card program, 24, 25
United Kingdom
appropriateness of treatment for coronary disease, 152, 165
coronary angiography and bypass surgery rates compared with United States, 32
tonsillectomy study, 26
University Hospitals Consortium
in-house technology assessment office, 138, 142
University of Pennsylvania
Leonard Davis Institute CEA guidelines development, 127
Upper gastrointestinal tract endoscopy
geographic variation in use, 32
Urinary incontinence guidelines, 151, 152-153, 170
UR. *See* Utilization reviews
U.S. Healthcare, Inc.
ABMT/HDC trial support, 94
report card project, 25
U.S. Preventive Services Task Force
clinical practice guidelines, 146, 150, 154, 156, 164, 167, 223-225
colorectal cancer screening, 166
*Guide to Clinical Preventive Services,* 146, *224,* 225
meta-analysis and, 64
ODPHP sponsorship of. 101
USPSTF. See U.S. Preventive Services Task Force
Utah
CQI study, 187, 188
Utilization reviews, 180, 190

**V**

VA. *See* Department of Veterans Affairs
Vaccinations. See *also* Immunizations
allowing nurses to administer, 180
CDC recommendations, 140
Vaccines
hepatitis B, 115
polio, 51
Vermont
clinical practice guidelines legislation, 35
tonsillectomy rates. 27
Vermont Trials Network, 89
Veterans Health Administration. *See also* Department of Veterans Affairs
activities, 101-102
Video disc technology for patient education, 181-182

**W**

Wade Park Veterans Affairs Medical Center
firms trial example, 58
Wash ington
back surgery PORT. 70

Wennberg and colleagues
benign prostatic hyperplasia treatment, 28-29
geographic variation in inpatient care utilization, 27
Women. See *also* Breast cancer; Cesarean delivery; Pregnancy
clinical trials registry of studies involving, 86-87
Pap screens for elderly women study, 111
Women's Health Study
assessment of serious or unknown side effects. 116
description, 56
Woolf
clinical practice guideline definition, 139

**Y**

Yusuf and colleagues
large, simple trial design, 54

**Z**

Zidovudine
AIDS treatment study, 83

☆ U.S. *GOVERNMENT* PRINTI NG OFFICE: 1994 - 301-804 - 814/27408