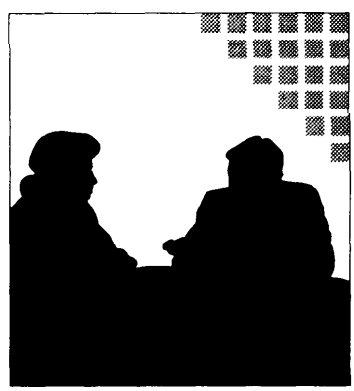
C Appendix C: Guideline Development Activities

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.