

Overview of Methodology for Lung Cancer Evidence Review and Guideline Development*

Douglas C. McCrory, MD, MHS; Gene L. Colice, MD, FCCP;
Sandra Zelman Lewis, PhD; W. Michael Alberts, MD, FCCP; and
Sydney Parker, PhD

A multidisciplinary panel was convened by the American College of Chest Physicians (ACCP) to develop clinical practice guidelines for lung cancer prevention, diagnosis, and treatment. The ACCP Expert Panel on Lung Cancer Guidelines produced 20 guidelines, each related to a distinct set of management decisions. This article describes the approach used to develop the guidelines, including identifying, evaluating, and synthesizing evidence, assessing the strength of evidence pertinent to individual guidelines, and grading guideline recommendations. (CHEST 2003; 123:3S–6S)

Key words: advisory committees; evidence-based medicine; lung neoplasms; practice guidelines

Abbreviations: ACCP = American College of Chest Physicians; NSCLC = non-small cell lung cancer; USPSTF = US Preventive Services Task Force

The American College of Chest Physicians (ACCP) Health and Science Policy Committee, through a poll of the ACCP membership and a structured internal review of all nominated topics, selected lung cancer as a high-priority topic for developing evidence-based clinical practice guidelines. The ACCP, with support from the Duke University Center for Clinical Health Policy Research, commissioned an expert panel, the goal of which was to produce clinically relevant, evidence-based guidelines for lung cancer prevention, diagnosis, and treatment that were aimed at primary care physicians and pulmonary specialists. Because the target audience is primary care physicians and pulmonologists, the guidelines are somewhat more detailed about issues of prevention, screening, and diagnostic evaluation compared to treatment issues. The guidelines avoid detailed recommendations about how to administer treatments, which usually are provided by other specialists, and instead focus

on recommendations for treatment approaches that are sufficient to make referral decisions.

The guideline topics are as follows: (1) prevention; (2) screening and early detection; (3) diagnosis; (4) initial evaluation; (5) physiologic and functional assessment; (6) solitary pulmonary nodules; (7) non-invasive staging; (8) invasive staging; (9) treatment of early-stage non-small cell lung cancer (NSCLC); (10) treatment of stage I NSCLC; (11) treatment of stage II NSCLC; (12) treatment of stage IIIA NSCLC; (13) treatment of stage IIIB NSCLC; (14) lung cancers with special considerations; (15) chemotherapeutic management of stage IV NSCLC; (16) small cell lung cancer evaluation and treatment; (17) follow-up/surveillance; (18) palliative treatment; (19) end-of-life care; and (20) practice organization. Not included among the topics is the management of other malignancies that also may occur in the lung, including mesothelioma, hamartoma, thymoma, and carcinoid and neuroendocrine tumors. Also omitted from these topics are complementary and alternative medicine treatments.

MATERIALS AND METHODS

Evidence Review

As a first step in identifying the evidence for each topic, we sought existing evidence syntheses including guidelines, systematic reviews, and meta-analyses. We searched computerized

*From the Department of Medicine (Dr. McCrory), Duke University Medical Center, Durham, NC; Pulmonary, Critical Care, and Respiratory Services (Dr. Colice), Washington Hospital Center, Washington, DC; the American College of Chest Physicians (Drs. Lewis and Parker), Northbrook, IL; and the H. Lee Moffitt Cancer Center and Research Institute (Dr. Alberts), Tampa, FL.

This research was supported by a contract from the American College of Chest Physicians.

Correspondence to: Douglas McCrory, MD, MHS, Duke Center for Clinical Health Policy Research, 2200 W Main St, Suite 220, Durham, NC 27705; e-mail: douglas.mccrory@duke.edu

bibliographic databases including MEDLINE, Cancerlit, CINAHL and HealthStar, the Cochrane Collaboration Database of Abstracts of Reviews of Effectiveness, the National Guideline Clearinghouse, and the National Cancer Institute Physician Data Query database. Computerized searches through July 2001 used the MeSH terms *lung neoplasms* (exploded) and *bronchial neoplasms* or text searches for lung cancer combined with review articles, practice guidelines, guidelines, and meta-analyses. We also searched and included studies from the reference lists of review articles, and queried experts in the field. An international search was conducted of Web sites of provider organizations that were likely to have developed guidelines. Abstracts of candidate English language articles were reviewed by two physicians (one with methodological expertise and one with content area expertise) and a subset was selected for review in full text. Full-text articles were reviewed again by two physicians to determine whether they were original publications of a synthesis and were pertinent to at least one of the topics of the guideline. Articles described as practice guidelines, systematic reviews, or meta-analyses were included, as were review articles that included a “Methods” section. Included articles were classified according to topic.

The methodological quality of existing clinical practice guidelines were evaluated using a rigorously developed and validated instrument (see the article on the critical appraisal of lung cancer guidelines in this supplement).¹

An initial review of published syntheses permitted the executive committee to select topics for new evidence reviews. Five topics for which existing syntheses were inadequate were selected for new evidence reviews conducted by the Duke Center for Clinical Health Policy Research. These topics included prevention, screening, diagnosis, and staging (invasive and noninvasive). The methods for these reviews are described in separate articles in this supplement. For other topics, identified guidelines, meta-analyses, and systematic reviews were made available to the writing committees described below.

Guideline Development

The guideline development panel was composed of members and nonmembers of the ACCP who were known to have expertise in various areas of lung cancer management and care, representing multiple specialties, and 13 national and international medical associations (Table 1), including the Alliance for Lung Cancer Advocacy, Support, and Education (a patient support group) and the Oncology Nurses Society. The specialties

included pulmonary/respiratory medicine, critical care, medical oncology, thoracic surgery, radiation oncology, epidemiology, law, and medical ethics. The panel was led by an executive committee including the chair (MA) and vice chair (GC), the leader of the methodology support group (DM), the ACCP project manager (SL), and the vice-president for Health and Science Policy (SP). The executive committee directed the guideline development process, methodological issues, panel composition, structure of the final document, and activities of the topic-specific writing committees.

Each writing committee was led by a member of the expert panel who agreed to review evidence and develop guidelines on 1 or more of the 20 topics listed earlier. The lead editor for each writing committee nominated individuals to assist with evidence review and guideline development, with the goal of including expertise from all relevant disciplines. Nominations for writing committee membership were reviewed and approved by the executive committee.

Funding for both the evidence reviews and guideline development was provided through an unrestricted educational grant from Bristol-Myers Squibb, which had no other role in the evidence review or guideline development process or content. Information about potential conflicts of interest were collected from each member of the expert panel or writing committee at the time of their nomination in accordance with the policy of the ACCP. Information on conflicts of interest for each panelist is listed in the guideline.

PROCESS

Each writing committee was charged with identifying the important management issues for which guidelines were needed. Each writing committee received a comprehensive list of existing systematic reviews and meta-analyses as well as guidelines published by other groups. In addition, for the five key topics described earlier, new systematic reviews were undertaken and were led by either the Duke Center for Clinical Health Policy Research or the Health Outcomes Research Group, Department of Epidemiology and Biostatistics, and Department of Medicine, Memorial Sloan-Kettering Cancer Center. For all other topics, writing committees were responsible for identifying and interpreting studies that were not otherwise covered in existing syntheses or guidelines.

To develop guidelines, the panel used informal group consensus techniques. The guidelines developed by the writing committees on each topic were distributed to the entire expert panel, and comments were solicited in advance of a meeting. During the meeting, proposed recommendations were reviewed, discussed, and voted on by the entire panel. Approval required consensus, which was defined as an overwhelming majority approval. Differences of opinion were accommodated by revising the proposed recommendation, the rationale, or the grade until consensus could be reached. The evidence supporting each recommendation was summarized, and recommendations were graded as described

Table 1—Organizations and Associations Represented by Guideline Development Panel Members

Panel Member Organizations
Alliance for Lung Cancer Advocacy, Support, and Education
American Association for Bronchology
American Cancer Society
American College of Physicians
American College of Surgeons Oncology Group
American Society of Clinical Oncology
American Society for Therapeutic Radiology and Oncology
American Thoracic Society
Association of Community Cancer Centers
Canadian Thoracic Society
National Comprehensive Cancer Network
Oncology Nurses Society
Society of Thoracic Surgeons

below. The assessments of level of evidence, net benefit, and grade of recommendation were reviewed by the executive committee.

VALUES

The panel considered data on functional status, quality and length of life, tolerability of treatment, and relief of symptoms in formulating guideline recommendations. Cost was not explicitly considered in the guideline development process. Data on these outcomes were informally weighted, without the use of explicit decision analysis or other modeling. The values placed on types of outcomes varied with clinical scenarios. For example, in some situations we considered life expectancy, such as the effects of early detection. In other situations we weighed quality of life more heavily, such as in palliative care and in interpreting small increases in life expectancy with chemotherapy for stage IV disease.

GRADING RECOMMENDATIONS

Recommendations were graded then using an adaptation of the scale used by the US Preventive Services Task Force (USPSTF), one of the most commonly

cited approaches (Table 2).² The USPSTF approach provided a framework for linking evidence to a conceptual model of how a clinical service affects health outcomes. Evidence is reviewed at the following three levels: (1) the individual studies; (2) the body of evidence concerning a single linkage in the conceptual model; and (3) the body of evidence relating interventions with health outcomes. At each level, the scheme offers general guidelines to assign one of the following three grades of evidence: good, fair, or poor. Good or fair quality evidence must include studies of sufficient design and quality to provide an unbroken chain of evidence-supported linkages that connect the intervention with health outcomes. In general, good evidence included prospective, controlled, randomized clinical trials, and poor evidence included case series and clinical experience. Trials with fair quality of evidence, for instance, historically controlled trials or retrospective analyses, were somewhere in between. In addition to the strength of the study design, however, study quality also was considered. The USPSTF approach considers well-recognized criteria in rating the quality of individual studies for a variety of different types of study design (eg, diagnostic accuracy studies and case-control studies). The thresholds for distinguishing good vs fair and fair vs poor evidence are not explicit but are

Table 2—Grades of Recommendation and Estimates of Net Benefit*

Grades	Description
Recommendation†	
A	The panel strongly recommends that clinicians routinely provide [the service] to eligible patients. An “A” recommendation indicates good evidence that [the service] improves important health outcomes and that benefits substantially outweigh harms.
B	The panel recommends that clinicians routinely provide [the service] to eligible patients. A “B” recommendation indicates at least fair evidence that [the service] improves important health outcomes and concludes that benefits outweigh harms.
C	The panel recommends that clinicians routinely provide [the service] to eligible patients. A “C” recommendation indicates that there was consensus among the panel to recommend [the service] but that the evidence that [the service] is effective is lacking, of poor quality, or conflicting, or the balance of benefits and harms cannot be reliably determined from available evidence.
D	The panel recommends against clinicians routinely providing [the service]. A “D” recommendation indicates at least fair evidence that [the service] is ineffective or that harm outweighs benefit.
I	The panel concludes that the evidence is insufficient to recommend for or against [the service]. An “I” recommendation indicates that evidence that [the service] is effective is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined, and that the panel lacked a consensus to recommend it.
Net benefit‡	
Substantial	Benefit greatly outweighs harm
Moderate	Benefit outweighs harm
Small/weak	Benefit outweighs harm to a minimally clinically important degree
None/negative	Harms equal or outweigh benefit, less than clinically important

*Adapted from Harris et al.²

†The grade of the strength of recommendations is based on both the quality of the evidence and the net benefit of the service (ie, test, procedure, etc.).

‡These levels of net benefit are based on clinical assessment. Estimated net benefit may be downgraded based on uncertainty in estimates of benefits and harms.

left to the judgment of panelists, reviewers, and members of the executive committee.

The grade of recommendation also depends on an assessment of the magnitude of benefit, harm, and net benefit, which is rated as substantial, moderate, small, or none/negative. Similar to the determination of quality of evidence, the determination of magnitude of benefit was a clinical judgment of the panel members (Table 2).

In assigning a grade to each recommendation, we first assigned a level of evidence, then estimated the magnitude of the net benefit, and used these two factors to provide the grade for the recommendation, which was coded as a letter (A, strongly recommended; to D, recommend against) [Table 3]. An “I” recommendation indicated situations in which the evidence is insufficient to determine net benefit. Our grading scheme is a modification of the USPSTF grades to allow recommendations for a

Table 3—Grade of Recommendation Based on Quality of Evidence and Estimate of Net Benefit

Quality of Evidence	Net Benefit			
	Substantial	Moderate	Small/weak	None/Negative
Good	A	B	C	D
Fair	B	B	C	D
Poor	C	C	C or I*	I

*Depending on whether panelists reached consensus or not, this grade could be C or I, respectively.

service when (1) evidence is poor, (2) the assessment of the net benefit is moderate to high, and (3) there is consensus among the expert panel to recommend it. This change was necessary because, unlike preventive services (*ie*, the routine offering of tests or treatments to well people) in which the burden of proof is high, clinical decisions about the treatment of patients with lung cancer often must be based on an interpretation of the available evidence, even if it is of poor quality. Our adaptation distinguished between interventions with poor evidence for which there is consensus (grade C) and interventions with poor evidence for which there is not consensus (grade I).

VALIDATION

After extensive review within the expert panel and executive committee, the guidelines were reviewed and approved by the ACCP Health and Science Policy Committee and then by the ACCP Board of Regents. The guidelines have not been field-tested.

REFERENCES

- 1 AGREE Collaborative Group. Appraisal of guidelines for research & evaluation (AGREE) instrument: second draft; August 2000. Available at: <http://www.agreecollaboration.org>. Accessed December 16, 2002
- 2 Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001; 20(suppl):21–35