MEDICARE COVERAGE ADVISORY COMMITTEE
OPERATIONS AND METHODOLOGY SUBCOMMITTEE

PROCESS FOR EVALUATION OF EFFECTIVENESS AND COMMITTEE OPERATIONS

Revised (Initial Meeting): July 21, 2005
Revised and approved (by CMS):   January 12, 2006

OVERVIEW

The Centers for Medicare & Medicaid Services (CMS) convened the Medicare Coverage Advisory Committee (MCAC) to provide advice on scientific, clinical practice, and ethical questions regarding Medicare coverage issues. The purpose of this document is to provide guidance to the MCAC for the conduct and product of its meetings. The goals of this document are to promote *consistency* in the reasoning that leads the MCAC to a conclusion about the scientific evidence and to facilitate *accountability* (to each other and to the public) by making that reasoning explicit. This document is designed to provide guidance on the manner in which MCAC should evaluate evidence and draw conclusions about how effectiveness should be evaluated.

The MCAC evaluation process consists primarily of two steps. First is an assessment of the *quality* of available evidence to draw conclusions about an intervention's effectiveness. Second is an evaluation of what the evidence demonstrates about effectiveness – that is, an evaluation of the *magnitude* of benefit conferred by the intervention.

At each of its meetings, the MCAC will use criteria and procedures to evaluate the quality of the scientific evidence and the magnitude of clinical benefit in determining the effectiveness of new medical products and services (laboratory test, diagnostic procedure, preventive intervention, treatment, and management) compared to standard alternatives. At the request of CMS or at the discretion of the Committee, the Committee may also provide advice about how to overcome shortcomings in the available evidence. The Committee may also discuss the likely consequences of technology dissemination on beneficiaries and the Medicare program.

This document has two purposes:

First, it provides general guidance to the committee in the form of suggestions about how to evaluate scientific evidence. This document emphasizes the distinction between quality of evidence and the magnitude of the benefit produced by a health intervention. The discussion is brief and at a general level. Background documents and references provide further discussion of methods for interpreting clinical evidence.

Second, it proposes specific procedures that the committee should follow in its deliberations. The purpose of these procedures is to ensure that the advice that MCAC provides to CMS is timely and meets the highest standards of comprehensiveness, balance, and scientific quality.

1

These principles and procedures should make the evaluation process as predictable, consistent, and understandable as possible. By making the reasoning behind the Committee's conclusions more explicit, these principles should also make the MCAC process more transparent.

The following recommendations are meant to assist the Committee in its deliberations until CMS issues further guidance. We will modify these recommendations as needed to respond to CMS guidance about the definition and application of the concept of "reasonable and necessary."

EVIDENCE FOR DELIBERATIONS

Overview: This process is intended to serve the public by identifying medical products and services that improve the health of Medicare beneficiaries. In advising CMS about the evidence that a new medical item or service is effective and likely to improve health outcomes of Medicare beneficiaries, the Committee will need to evaluate the quality of evidence and to draw conclusions about the implications of the evidence.

The Committee should explore many sources in assembling the body of evidence to be used in their deliberations. The sources of evidence might include the peer-reviewed scientific literature, the recommendations of expert committees, and unpublished data used to secure FDA approval. The quality of the evidence from these sources will vary, and the committee should weigh the evidence according to its quality.

Outcomes evaluated: The Committee should consider several health outcomes as part of its deliberations. The committee should rate how, compared to alternative or standard management approaches for the condition under review, the intervention affects:

- Quality of Life
- Morbidity
- Mortality
- Diagnostic Accuracy (for diagnostic interventions) and impact on management (See Appendix)
- Other health outcomes as appropriate, such as re-hospitalizations

The MCAC greatly values information on the effect of treatments on quality of life, functional status and other relevant aspects of health. In keeping with the MCAC's scientific mission, the most valuable data regarding these outcomes are those collected as part of a scientific study, such as a clinical trial. Other information can complement the results of formal studies. For example, patients experiencing a health condition can provide insight into types of benefits and risks that may have been overlooked or poorly measured in existing studies. They can, thereby, aid in the evaluation of the current evidence base and help direct future research. Patient testimonials, like other information relevant to the Committee's deliberations, should be provided to committee members in writing in advance of the meeting, so that it can be receive full and thoughtful consideration. However, information provided by direct patient commentary should not be considered a replacement for information derived from rigorous studies. Testimony that an intervention was or was not effective in a single patient's experience ordinarily provides little

benefit to the Committee's task, since the attribution of the health outcome to a given intervention in a single case typically cannot be made with confidence.

Quality of evidence: **The major role of the Committee is to determine whether the scientific evidence is of adequate quality to draw conclusions about the effectiveness of the intervention in routine clinical use in the population of Medicare beneficiaries.**

Assessing the adequacy of the quality of the evidence is a *sine qua non* of essentially all modern approaches to the evaluation of medical technologies. Defining the criteria for adequate quality of evidence is a critical step. The committee's definition of adequate evidence includes both the *validity* of the evidence and its *general applicability* to the population of interest, i.e. *generalizability*.

Many forms of evidence can be valid or not, depending on circumstances specific to the individual study. The most rigorous type of evidence is derived from randomized controlled clinical trials. The ideal randomized clinical trial has appropriate endpoints established before the trial starts, enrolls a representative sample of patients, is conducted in clinical practice in the patient population of interest, and evaluates interventions (diagnostic tests, surgical procedures, medical devices, drugs) as typically used in routine clinical practice.

When several such well-designed trials yield consistent results, there is likely to be a strong consensus that the evidence is sufficient. However, this level of evidence will be unavailable for many of the interventions that the MCAC will evaluate. There may be randomized trials conducted in other populations (e.g., middle-aged men rather than men and women 65 years of age and older), randomized trials with important design flaws (e.g., not double-blinded), or non-randomized studies with concurrent or historic controls. Deciding whether such studies constitute valid, generalizable evidence can be very difficult.

The Committee believes that general guidelines for deciding whether the evidence is adequate will serve its purposes better than a rigid set of standards. In considering the evidence from any study, the MCAC should try to answer two main questions:

a) How close are the effects measured in the study to their true value(s)?

The degree to which the study result differs from the underlying truth is composed of two factors; *chance* and *bias*. The confidence interval (CI) around the estimated effect is intended to capture the role of *chance*; it measures the underlying range of true effects that are compatible with the estimated value. The confidence interval is critical in deciding whether a study has statistically "ruled out" either a zero (null) effect, or, in the case of non-significant results, a clinically important effect. It is critical to recognize that a statistically significant effect may not be clinically important or meaningful in any other way. Conversely, statistically non-significant estimates may not rule out important effects and may collectively provide strong evidence against the null hypothesis.

3

Other errors of inference can arise from fundamental flaws in the study design or analysis, rather than pure chance. These errors are said to be the result of *bias*. An estimate of effectiveness, or any other number that a study is designed to measure, is said to be *unbiased* if its average or expected value is equal to its true value. An estimate is biased, therefore, if repeating the study that generated it would, on average, produce an estimate that also deviates from the truth.

In medical studies, as in most other contexts, bias frequently occurs because measured or unmeasured patient characteristics affect the outcome that is being measured, and the effects of these characteristics are falsely attributed to the intervention. For example, observational studies of the effects of high-dose chemotherapy on the survival of women with breast cancer suggested that the therapy provided a substantial benefit when compared to conventional chemotherapy. Randomized trials eventually showed, however, that high-dose chemotherapy conferred no survival benefit. Among the explanations for these findings is that women who had poor health (who would have higher mortality rates regardless of the mode of therapy) were not usually offered the more demanding high-dose regimen. Thus, the effects of their underlying health were falsely attributed to conventional chemotherapy. Bias from this cause is frequently described as *selection bias* – the patients selected to receive an intervention are either healthier or less healthy on average than those who receive an alternative intervention in ways that are not fully reflected in the characteristics measured as part of the study.

Randomized trials are viewed as the best approach to avoiding bias because randomization ensures that, on average, measured and unmeasured characteristics are the same for study subjects assigned to each arm of the trial. Randomization increases confidence that the expressions of uncertainty (e.g., confidence intervals) about a trial's estimates of effect size (and other measured outcomes) are correct. The effects of uncertainty due to random variation diminish as sample size increases. Although there can be random variation in the characteristics of patients assigned to the different arms of a randomized trial, any differences in underlying health should not differ systematically.

Though bias can sometimes be minimized or eliminated through analytic means, it is often not correctable. An important task of the MCAC will be to assess whether the study design is likely to lead to bias, and if so, to consider how large the bias is likely to be. The Committee should also consider the magnitude of uncertainty due to simple chance variation, drawing conclusions about the range of effect sizes that are consistent with the experimental evidence.

b) How applicable are the results to the Medicare population, in the settings in which they receive care?

The studies reviewed by the MCAC are often conducted in settings that differ from those in which the typical Medicare beneficiary receives care. Many studies of new procedures are conducted in academic medical centers and other institutions that provide a high volume of specialized care and offer a broad set of services. Neither the specific details of the procedures nor the outcomes that result may be comparable in the diverse institutional and community settings in which most Medicare beneficiaries receive care. Furthermore, the patients enrolled in trials and other studies may be neither elderly nor disabled, leading to doubts that the

4

interventions would produce the same results in the Medicare population. A key task of the MCAC is determine whether the results reported in studies are likely to apply to Medicare beneficiaries in the settings in which they receive care.

Generalization to populations not included in the studies, and to settings of care that are dissimilar to those included in the studies, is inherently difficult. There is no single method or set of methods that will be valid in all circumstances, and available data may be too limited to draw firm conclusions about whether the results of efficacy studies apply in different settings or in different populations. Nevertheless, the Committee should seek to draw upon the best methods and sources of information to address applicability of the studies. For example, although the MCAC is likely to weigh randomized trials and observational studies heavily when they evaluate a clinical intervention, they may use other data and apply techniques such as decision analytic and epidemiological modeling in order to estimate effects in the Medicare population.

Such models can be very helpful in assessing the effects of an intervention in a population excluded from published studies. Often the chief question is whether an intervention that reduces the risk of death or some other serious adverse event in a study population will have similar effects in a population that is at a much lower or higher risk of death. Any intervention that cuts the risk of death from heart attack in half will have much larger effects on mortality in a group of survivors of myocardial infarction who have a 5% annual risk of cardiac death than in a group of asymptomatic young men, who may have a 0.5% annual risk of cardiac death. A common approach to extrapolation is to assume that the *relative risk reduction* from a health intervention is common across populations. If the assumption is valid, the relative risk reduction from a trial can be applied to a population that has a different risk of the outcome in question (e.g., mortality). In the cardiac death example, the intervention would cut the risk of cardiac death by 2.5% (from 5% to 2.5% in absolute terms) in the first group (which is the one likely to have been included in a study) and by 0.25% (from 0.5% to 0.25%) in the second group. This simple assumption thus makes it possible to estimate the change in the absolute risk that would result from use of the intervention in a population that was not adequately represented in the trial. Of course, this approach would not be valid if the group that received a greater absolute benefit was also substantially more or less likely to experience adverse effects from the intervention. The validity of this assumption, and any other assumptions used in such models, will need to be assessed in the context of the specific intervention under consideration.

Extrapolation to other settings of care can be particularly difficult. The Committee may wish to compare, for example, complication rates resulting from the use of a surgical procedure in routine settings to the rates reported in trials and other studies. For a novel procedure that has not yet been used outside experimental settings, however, comparative information will not be available. In such cases, the MCAC may be asked to discuss whether similar outcomes are likely to be obtained in routine practice settings. In answering this question, they should consider the skill and training - of both physicians and support staff - required to provide the intervention. They should discuss whether the manner in which the treatment is likely to be applied in practice corresponds to the treatment delivered under the experimental protocol. Finally, they should consider what further information would be most helpful in addressing these questions.

Size of Health Effect and net health outcomes: Evidence from well-designed studies must establish how the effectiveness of the new intervention compares to the effectiveness of established services and medical interventions. If the evidence is adequate to draw conclusions (as defined above) about the magnitude of the effect, the next question is the clinical importance of the size of the effect compared with interventions that are widely used, and whether there is a net health benefit, i.e., does the magnitude of beneficial health effect outweigh the adverse health effects. This judgment should take into account both the size of these effects and the serious of the related outcomes. The Committee should help CMS make coverage decisions by discussing the size and direction of the intervention's effect on all health outcomes as compared to other standard interventions.

THE CASE OF INSUFFICIENT EVIDENCE

CMS may ask the MCAC for advice when the evidence for effectiveness or safety is ambiguous, scanty, or of poor quality. When an MCAC determines that the evidence is insufficient to draw conclusions about the effectiveness of an intervention, it will not attempt to assess or discuss the net health outcomes. Instead, it will explain the reason for its determination and also form a judgment about:

- the possibility of developing better evidence
- the potential consequences of waiting to obtain better information or of permitting dissemination with insufficient knowledge of effects
- patient and caregiver views

CMS could deal with the problem of inadequately studied but promising technologies in several ways:

- It might encourage studies that would provide adequate evidence about the effectiveness of promising technologies by directly supporting research.
- It could approve coverage under a clinical trial or similar protocol. For example, the technology would be covered only when it is used in the context of an approved, well-designed study that is likely to fill the important gaps in the evidence. This should accomplish the desired goal of making the device available while assuring that a body of evidence is collected to facilitate a definitive coverage determination (See Guidance Document on Coverage with Evidence Development).
- It could make a coverage decision based upon the best interpretation of the available evidence. Such an approach would give CMS the flexibility to cover promising treatments for conditions that are too rare to support definitive study.

COMMITTEE OPERATIONS

1. **Explanation: The Committee must explain its recommendations to CMS in writing, either through meeting minutes or other summaries.**

Adherence to this principle will help ensure the integrity of the MCAC procedures and judgments and, by making the Committee's reasoning processes more explicit and open, provide internal and external accountability. The requirement for written explanations will also help to diminish the risk of ambiguity and misunderstanding between the MCAC, CMS, and the public.

2. **Structure of evidence provided to the committee: The Committee should receive well-organized, high quality relevant background information before beginning its deliberations. The evidence should be summarized in a report, not simply presented as a collection of data or primary studies.**

The integrity of the coverage decision process begins with a complete critical review of the evidence, which will be summarized in a document entitled the Evidence Report.

The MCAC has identified several issues related to the Evidence Report:

The standard of excellence for the Evidence Report: The standard of excellence should be the best work in the private sector (e.g., the Blue Cross-Blue Shield Association), by professional organizations (e.g., the American College of Physicians), and for other federally sponsored committees (e.g., the Evidence-based Practice Centers technical support for the U.S. Preventive Services Task Force). The evidence reports to be used in MCAC deliberations should provide a comprehensive review and summary of the state of the science surrounding the issue before the MCAC meets. The production of a full Evidence Report on a typical MCAC topic should seldom require more than six months after CMS has received a request for consideration of reimbursement for an intervention (see attached slide).

Formulation of the key questions for the Evidence Report to address: The value of an Evidence Report will depend in large part upon the questions that it seeks to answer. CMS staff and MCAC chair and vice-chair will draft the key questions.   The evidence report must address these key questions for optimal operations of the Committee.

Dissemination of the Evidence Report: The Evidence Report should become a means to promote effective dialog at panel meetings. To this end, CMS should post the Evidence Report on the MCAC website and notify interested parties at least one week before a Committee meeting.

3. **Committee member involvement: Committee members should take an active role in reviewing the evidence. The Committee chair should play an active role in framing the questions that the evidence report must address and the Committee must answer.**

The MCAC should include some people who have acquired expertise in the topic of the coverage request to assure that the Committee can fairly evaluate the evidence review.

Each Committee member should read the evidence report carefully and understand the main issues that the report addresses.

SUGGESTED READING

JAMA Users' Guide to the Medical Literature

The following articles were originally published in the Journal of the American Medical Association. They were written for clinicians and emphasize the basics of interpreting medical literature to make patient care decisions. They have been collected in the book *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* by Gordon H. Guyatt and Drummond Rennie, Chicago: American Medical Association, 2001.

 They are also available on the web at
   http://www.cche.net/usersguides/main.asp .

Guyatt G, Rennie D and the Evidence Based Medicine Working Group. Why Users' Guides? EBM Working Paper Series #1. Only available on the Internet as:
http://www.cche.net/principles/content_why.asp

Guyatt GH. Users' guides to the medical literature [editorial]. *JAMA* 1993; 270 (17) : 2096-2097.

Oxman A, Sackett, DL & Guyatt GH. Users' guides to the medical literature. I. How to get started. *JAMA* 1993 Nov 3; 270 (17) : 2093-2095.

Guyatt GH, Sackett DL and Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993; 270 2598-2601.

Guyatt GH, Sackett DL and Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994; 271:59-63.

Jaeschke R, Guyatt G and Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA* 1994 Feb 2; 271 (5) : 389-391.

Jaeschke R, Gordon H, Guyatt G & Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. what are the results and will they help me in caring for my patients? *JAMA* 1994; 271 : 703-707.

Levine M, Walter S, Lee H, Haines T, Holbrook A & Moyer V. Users'guides to the medical literature. IV. How to use an article about harm. *JAMA* 1994 May 25; 271 (20) 1615-1619.

Laupacis A, Wells G, Richardson S & Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994; 272 : 234-237.

Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA*. 1994 Nov 2;272(17):1367-71.

Richardson WS , Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis A. Are the results of the study valid? *JAMA* 1995; 273 (16) : 1292-1295.

Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis B. What are the results and will they help me in caring for my patients? *JAMA* 1995; 273 (20) : 1610-1613.

Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines A. Are the recommendations valid? *JAMA* 1995; 274 (7) : 570-574.

Wilson MC, Hayward RSA, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines B. What are the recommendations and will they help you in caring for your patients? *JAMA* 1995 Nov 22-29; 274 (20) : 1630-1632.

Guyatt GH, Sackett DL, Sinclair JC et al. Users' Guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995 Dec 13; 274 (22) : 1800-1804.

Naylor CD and Guyatt GH Users guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. Evidence-Based Medicine Working Group. *JAMA* 1996 Feb 21; 275(7) : 554-558.

Naylor CD and Guyatt GH. Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. Evidence-Based Medicine Working Group. *JAMA* 1996 May 8; 275 (18) : 1435-1439.

Guyatt GH, Naylor CD, Juniper E et al. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. *JAMA* 1997 Apr 16; 277 (15) : 1232-1237.

Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1997 May 21;277(19):1552-1557.

O'Brien BJ. Heyland D. Richardson WS. Levine M. Drummond MF. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group [published erratum appears in JAMA 1997 Oct 1;278(13):1064]. [Journal Article] *JAMA* 277(22):1802-6.

Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the Medical Literature. XIV. How to decide on the applicability of clinical trial results to your patient. Evidence Based Medicine Working Group. *JAMA* 1998; 279 (7): 545-549.

Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis.*JAMA* 1999 Apr 7;281(13):1214-1219.

Guyatt GH, Sinclair J, Cook DJ, Glasziou, P Users' guides to the medical literature: XVI. How to use a treatment recommendation. *JAMA* 1999 May 19;281(19):1836-1843.

Barratt A, Irwig L, Glasziou P, et.al.   Users guide to medical literature XVII How to use guidelines and recommendations about screening. *JAMA* 1999;281:2029

Randolph AG, Haynes RB, Wyatt JC, Cook DJ, Guyatt GH. Users guide to medical literature. XVIII How to use an article evaluating the clinical impact of a computer-based clinical decision support system. *JAMA* 1999;282: 67-74:

Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. *JAMA* 1999 Aug 25;282(8):771-8

McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' Guides to the Medical Literature: XIX. Applying clinical trial results. B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA*. 1999 Oct 13;282(14):1371-7.

McAlister FA, Straus SE, Guyatt GH, Haynes RB. Users' guides to the medical literature: XX. Integrating research   evidence with the care of the individual patient. JAMA 2000 June 7;283(21):2829-2836

Hunt DL, Jaeschke R, McKibbon KA. Users' guides to the medical literature: XXI. Using electronic health information resources in evidence-based practice. Evidence-Based Medicine Working Group. *JAMA*. 2000 Apr 12;283(14):1875-9.

McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII:
How to use articles about clinical decision rules. JAMA 2000 Jul 5;284(1):79-84

Giacomini MK, Cook DJ. Users' guides to the medical literature XXIII. Qualitative research in health care A. Are the
Results of the Study Valid? JAMA. 2000 Jul 19;284(3):357-362

Giacomini MK, Cook DJ. Users' guides to the medical literature XXIII. Qualitative research in health care B. What are the results and how do they help me care for my patients? JAMA 2000 Jul 26;284(4):478-482

Richardson WS, Wilson MC, Williams JW, Moyer VA, Naylor CD. Users' guides to the medical literature XXIV. How to use an article on the clinical manifestations of disease JAMA 2000 Aug 16;284(7):869-875

Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC. Users' guides to the medical literature: XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. JAMA 2000 Sep 3; 284 (10): 1290-1296.

United States Preventive Services Task Force Methodology

The methods of the U.S. Preventive Services Task Force have strongly influenced the approach to rating the quality of clinical evidence and applying it to decisions about adoption of preventive services. The methods are described in the article,

Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, for the Methods Word Group, third U.S. Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20(3S):21-35. The article is available at http://www.ahrq.gov/clinic/ajpmsuppl/harris1.htm.

APPENDIX: Guidelines for Evaluating Diagnostic Tests

When asked to evaluate diagnostic tests, the MCAC can apply criteria that are similar to those used for other health interventions that come before the Committee. The MCAC will need to determine whether the evidence is adequate to conclude that the diagnostic test improves outcomes and, if the evidence is adequate, to classify the magnitude of the health benefit when a test is used for a specific purpose. When good quality studies directly measure how the use of a diagnostic test affects morbidity, mortality, and other health outcomes, the Committee can easily determine that the evidence is adequate and draw conclusions about the magnitude of the health benefits. Then the evaluation will be essentially identical to that of a therapeutic intervention.

However, direct proof of effectiveness of diagnostic tests is usually unavailable. Typical studies that evaluate the effectiveness of tests focus either on technical characteristics (e.g., does a new imaging modality produce higher resolution images) or effects on accuracy (does it distinguish between patients with and without a disease better than another test). The MCAC can sometimes, but not always, draw conclusions about the effectiveness of a test from such information.

When the best studies measure only the accuracy of the test, the Committee will have to determine whether the evidence is adequate to conclude that the test improves the accuracy of diagnosis or staging of disease *and* that the improvement in accuracy leads to better health outcomes compared to those from an alternative clinical strategy. The alternative strategy could be, for example, the use of another test, use in combination with another test, or the use of no test at all (e.g., the alternative is treatment or observation without testing). If *direct* evidence linking the use of the test to health outcomes is not available, the Committee should answer the following questions, which collectively determine whether there is convincing *indirect* evidence that the test will lead to better health outcomes:

Question 1: *Is the evidence adequate to determine whether the test provides more accurate diagnostic information?*

Question 1 applies when the alternative under consideration is another diagnostic strategy. The definition of "more accurate" is crucial. The standard measures of accuracy are **sensitivity** (probability of a positive test result in a patient with a disease or risk factor or other health condition) and **specificity** (the probability of a negative test result in a patient who does not have the disease). Ideally a new test would increase *both* sensitivity and specificity, but often it does not because in most clinical situations there is a tradeoff between increased sensitivity and increased specificity. For decision making purposes tests used to *rule out* the presence of disease must have high sensitivity. Conversely, ruling in disease requires tests of high specificity. In deciding whether one test is more accurate than a second, established test, the Committee will need to evaluate the quality of the studies of test performance. If the two tests are to be used together, then the committee will need to evaluate the quality of the studies of both tests as well as their *combined* impact on patient management.

The design and evaluation of trials related to diagnostic tests tends to be more difficult than those for therapies because the "gold standard" of truth is frequently missing in these trials or the source of patients may be unrepresentative (e.g., by stage of disease or site of care). The MCAC will specifically need to decide whether the estimated accuracy of a test in a study is likely to be distorted by a substantial degree of bias or whether the limitations of the study are sufficiently minor that it is possible to draw conclusions about the accuracy of the test.

Question 2: *If the test changes accuracy, is the evidence adequate to determine how the changed accuracy affects health outcomes?*

To determine whether a difference in test accuracy would lead to important changes in health outcomes, it is necessary to combine multiple sources of information. Among the information required is prevalence of disease in the tested population, the probabilities of positive and negative test results, the actions that would be taken in response to the test, and the consequences of those actions for health.

Drawing such inferences requires a great deal of information beyond basic test performance characteristics. The benefits of testing are largest and therefore easiest to estimate when the treatment or management strategy is effective for patients with the disease but poses risks or discomfort that would not be acceptable when administered to patients who do not have the disease. Then, improved accuracy leads to effective treatment for more people who truly have the disease, while helping to avoid unnecessary treatment in people who would not benefit from it.

Thus, although the evidence that diagnostic tests for cancer and for heart disease alter health outcomes is largely indirect, it is often compelling. For these categories of disease, there is often strong evidence that treatments with significant adverse consequences are effective when used appropriately. Committee will need to judge whether the test leads to better patient management by increasing the rate at which patients with disease receive appropriate treatment while reducing the rate at which patients who do not have the disease receive unnecessary treatment.

If management changes, the improvement in health outcomes should be large enough to convince the Committee that it is clinically significant. A small increase in accuracy can lead to substantial improvements in health outcomes if treatment is highly effective. Improved accuracy is of little consequence, however, if treatment is either ineffective, so there is little benefit to patients with the disease, or very safe, so there is little harm to patients without the disease. When a treatment has little effect on anyone, improved accuracy is unlikely to lead to improved health outcomes or even to influence clinical decisions.

Under exceptional circumstances, prognostic information, even if it did not affect a treatment decision, could improve health outcomes by improving a patient's sense of well-being. The Committee should be alert for circumstances in which patients would be likely to value prognostic information so much that the information would significantly alter their well-being.

Decision analytic modeling can be a particularly useful tool for integrating data from multiple sources to make estimates of the potential effects of diagnostic testing on health outcomes. However, when the results from such models are used by the MCAC to reach conclusions about health outcomes, model assumptions, parameter estimates, and results should be provided in sufficient detail for the Committee to judge the quality of the studies.

Summary

The recommended approach for evaluating diagnostic tests is, therefore, as follows:

- Review, when available, high quality studies that provide *direct* evidence that test results improve health outcomes.
- If there is no high quality *direct* evidence, determine the extent to which there are changes in patient management, particularly when the management strategy is effective in patients with the disease and does not benefit or even harms those without the disease.