

SPECIAL ARTICLE

Tumor Marker Utility Grading System: a Framework to Evaluate Clinical Utility of Tumor Markers

*Daniel F. Hayes, Robert C. Bast, Christopher E. Desch, Herbert Fritsche, Jr., Nancy E. Kemeny, J. Milburn Jessup, Gershon Y. Locker, John S. Macdonald, Robert G. Mennel, Larry Norton, Peter Ravdin, Sheila Taube, Rodger J. Winn**

Introduction of tumor markers into routine clinical practice has been poorly controlled, with few criteria or guidelines as to how such markers should be used. We propose a Tumor Marker Utility Grading System (TMUGS) to evaluate the clinical utility of tumor markers and to establish an investigational agenda for evaluation of new tumor markers. A Tumor Marker Utility Grading Worksheet has been designed. The initial portion of this worksheet is used to clarify the precise characteristics of the marker in question. These characteristics include the marker designation, the molecule and/or substance and the relevant alteration from normalcy, the assay format and reagents, the specimen type, and the neoplastic disease for which the marker is being evaluated. To determine the clinical utility of each marker, one of several potential uses must be designated, including risk assessment, screening, differential diagnosis, prognosis, and monitoring clinical course. For each of these uses, associations between marker assay results and expected biologic process and end points must be determined. However, knowledge of tumor marker data should contribute to a decision in practice that results in a more favorable clinical outcome for the patient, including increased overall survival, increased disease-free survival, improvement in quality of life, or reduction in cost of care. Semiquantitative utility scales have been developed for each end point. The only markers recommended for use in routine clinical practice are those that are assigned utility scores of “++” or “+++” on a 6-point scale (ranging from 0 to +++) in the categories relative to more favorable clinical outcomes. Each utility score assignment should be supported by documentation of the level of evidence used to evaluate the marker. TMUGS will establish a standardized analytic technique to evaluate clinical utility of known and future tumor markers. It should result in improved patient outcomes and more cost-efficient investigation and application of tumor markers. [J Natl Cancer Inst 1996;88:1456-66]

Tumor markers are ostensibly used in the clinical setting to provide information that will influence clinical decision-making (1). However, unlike the objective criteria established to evaluate new therapeutic agents, few guidelines have been established to determine if and/or when use of a tumor marker should become standard (2). For example, although trial designs for evaluation of new therapeutic agents are not identical, there is general consensus regarding the framework of phase I, II, and III trials and regarding the use of various semiquantitative scales (response criteria, toxicity scales, performance criteria scales) to measure outcomes within these studies. There is also general consensus regarding the relative merits of prospective clinical trials versus retrospective reviews of clinical experience. Nonetheless, each of these trial design frameworks is sufficiently flexible to permit individual investigator initiatives and modifications. Within this framework, most investigators and clinicians understand and use common terms, such as complete or partial response, or such as grade I-IV toxicity, to determine how and when a new agent is ready to be accepted as “standard.” We propose that it is appropriate to establish similar criteria for evaluation of tumor markers and to standardize the tumor marker information for clinical utility.

**Affiliations of authors:* D. F. Hayes, Breast Evaluation Center, Dana-Farber Cancer Institute, Boston, MA; R. C. Bast, H. Fritsche, Jr., R. J. Winn, The University of Texas M. D. Anderson Cancer Center, Houston; C. E. Desch, Massey Cancer Center, Medical College of Virginia, Richmond; N. E. Kemeny, L. Norton, Memorial Sloan-Kettering Cancer Center, New York, NY; J. M. Jessup, New England Deaconess Hospital, Boston; G. Y. Locker, Evanston Hospital, IL; J. S. Macdonald, Temple University Cancer Center, Philadelphia, PA; R. G. Mennel, Texas Oncology, Dallas; P. Ravdin, University of Texas Health Science Center at San Antonio; S. Taube, Cancer Diagnosis Branch, Division of Cancer Biology, Diagnosis, and Centers, National Cancer Institute, Bethesda, MD.

Correspondence to present address: Daniel F. Hayes, M.D., Breast Cancer Program, Lombardi Cancer Center, Rm. E504, Research Bldg., 3970 Reservoir Rd., N.W., Washington, DC 20007.

See “Notes” section following “References.”

Potential difficulties in using surrogate markers for predicting clinical end points have been discussed previously (3,4). For example, if tumor marker data are unreliable or if assumptions regarding the utility of tumor marker data are incorrect, clinical decision-making will be adversely affected. A patient's treatment plan might be altered on the basis of tumor marker data without evidence that such an action is justified. It is generally accepted that inappropriate administration of drugs may well result in a poor outcome. Likewise, improper clinical decisions based on incorrectly interpreted tumor marker results may not only increase cost of care but also may expose the patient to adverse consequences, such as treatment with toxic but unnecessary therapeutic agents. To aid in the appropriate use of tumor markers, we propose a system for tumor marker evaluation that will allow objective assessment for acceptance or rejection of a marker for use in clinical practice, in a manner analogous to the systems already in use for new drug development.

We have developed a standardized method of defining a specific tumor marker to be evaluated, as well as a Tumor Marker Utility Grading System (TMUGS). Key features of TMUGS include proposing semiquantitative utility scales and establishing a hierarchy of levels of evidence to support conclusions regarding the utility of a given marker for a given use. We have generated a single page "worksheet" into which all of the features of the TMUGS are entered (Fig. 1). The user may

wish to extract portions of the Tumor Marker Utility Grading (TMUG) Worksheet, depending on the intended function.

Who might use the TMUGS? This system was expressly designed for the purpose of practice guideline development (see "Notes" section). Therefore, we suggest that the TMUGS will aid clinicians in the determination of whether currently available tumor markers are appropriately used in practice. However, we also suggest that clinical and laboratory investigators might consider a modified use of the TMUGS as they plan their studies of new tumor markers and as they provide expert review of investigations of others. Designing studies within the TMUGS framework should lead to more efficient determination of the clinical utility of a new marker.

It is not the purpose of this article to evaluate any specific marker or analytical technique (biochemical or statistical) for any particular use. Rather, it is to introduce a consistent and objective process for evaluating tumor markers. We propose that the TMUGS is a step toward helping to standardize and establish some order in the presently chaotic field of tumor markers.

Definitions and Specifications of Tumor Markers

Various designations are often used interchangeably for tumor markers. By definition, a marker represents a qualitative or quantitative alteration or deviation from normal of a mole-

| DEFINITIONS AND SPECIFICATIONS OF MARKER | | | | | | | | | | | | |
|---|--|--|---|--|--|--|---|---|--|--|--|--|
| MARKER DESIGNATION/ NOMENCLATURE (e.g. ER, P53, CEA, etc.) | | | MOLECULE/SUBSTANCE ASSAYED & ALTERATION DETECTED (e.g. DNA/mutation, RNA/overexpression, Protein/increased levels, etc.) | | | | ASSAY FORMAT TO DETECT MARKER (e.g. EIA, ICA, SSCP, etc.) | | | | | |
| REAGENTS USED (e.g. specific MAb or probe, or commercial assay) | | | | | | | | SPECIMEN SOURCE (e.g. frozen or fixed tissue, plasma, urine, circulating cells, etc.) | | | | |
| DISEASE (e.g. breast cancer, colon cancer, etc.) | | | | | | | | | | | | |

| UTILITY | Marker association with biologic: | | | | Use leads to decision in practice that results in a more favorable clinical outcome: | | | | | | | |
|--|-----------------------------------|-------------------|---------------|-------------------|--|-------------------|-----------------------|-------------------|-----------------|-------------------|---------------|-------------------|
| | Process | | Endpoint | | Survival | | Disease Free Survival | | Quality of Life | | Cost of Care | |
| | Utility Score | Level of Evidence | Utility Score | Level of Evidence | Utility Score | Level of Evidence | Utility Score | Level of Evidence | Utility Score | Level of Evidence | Utility Score | Level of Evidence |
| Uses | | | | | | | | | | | | |
| Determine risk | | | | | | | | | | | | |
| Screening | | | | | | | | | | | | |
| Differential diagnosis | | | | | | | | | | | | |
| Prognosis: Predict relapse/progression | | | | | | | | | | | | |
| Primary | | | | | | | | | | | | |
| Metastatic | | | | | | | | | | | | |
| Prognosis: Predict response to therapy | | | | | | | | | | | | |
| Primary | | | | | | | | | | | | |
| Metastatic | | | | | | | | | | | | |
| Monitor course | | | | | | | | | | | | |
| Detect relapse in patient with no evidence of disease after therapy for primary or recurrent disease | | | | | | | | | | | | |
| Follow detectable disease | | | | | | | | | | | | |

Fig. 1. Tumor Marker Utility Grading Worksheet. ER = estrogen receptor; CEA = carcinoembryonic antigen; FNA = fine-needle aspiration; EIA = enzyme-linked immunoassay; ICA = immunocytochemical assay; SSCP = single-strand conformational polymorphism; MAb = monoclonal antibody.

cule, substance, or process that can be detected by some type of assay. For the purposes of this proposal, only those alterations that are detected by a special assay performed on a body tissue or fluid, other than routine histopathologic or laboratory evaluations, will be considered as tumor markers.

The TMUG Worksheet (Fig. 1) is designed to organize multiple areas of heterogeneity and confusion that have contributed to the lack of a standardized approach to clinical acceptance of specific tumor markers. The top portion of the worksheet includes precise designations and nomenclatures that specifically identify the marker of interest. The bottom portion provides a framework in which one can evaluate the available data for that specific tumor marker in regard to one of several different clinical uses and one of several different outcomes. Succeeding tables provide more detailed explanations of one or more categories within the worksheet.

Marker Designations and Nomenclatures

For each tumor marker evaluation, the marker in question must be precisely designated. Many designations may be given to the same family of markers, and one designation may be given to many different types of markers (Fig. 1, Table 1). Markers may acquire different designations when different investigators or commercial interests independently identify and develop assays for potentially useful markers. For example, the HER-2 (for human epithelial receptor-2) gene is also known as neu (for neuroblastoma oncogene) and erb-B2 (for epithelial receptor b-B2) (5).

Molecule or Substance Assayed and Alteration Detected

Gene designation may have different meanings, depending on what is being assayed. For example, abnormalities in the HER-2/neu/erb-B2 axis may describe genetic, RNA, or protein changes in tissue. These changes include mutations, amplifications, or overexpression (6). Another abnormality in the HER-2/neu/erb-B2 axis includes elevations of circulating levels of the external domain of the proto-oncogene protein (7-9). Furthermore, a report (10) has described evidence of an immune response against HER-2/neu, as reflected by endogenous anti-

body titers or cellular immune activity directed toward the protein. Finally, although not yet described, identification of abnormalities in one or more HER-2/neu/erb-B2 ligands may also be useful as tumor markers (11). Therefore, in this example, the statement that "HER-2/neu is positive" may have various implications. These ambiguities make it essential to designate precisely what molecule or process are altered (e.g., DNA, RNA, protein, antibody, cellular response, etc.) and what alteration was detected in that molecule or process (e.g., amplification, mutation, deletion, overexpression, elevated soluble protein levels, increased cellular activity, etc.) (Fig. 1, Table 1). These may be quite specific. For example, one mutation in a tumor suppressor gene such as p53 may not produce the same biologic effect as another mutation or a deletion in the same gene (12).

Assay to Detect Alteration in Molecule or Substance of Interest

An assay is a test to identify the alteration in the marker substance or process. One cannot assume that two assays for the same alteration of the same molecule provide identical results. Rather, results can be expected to be quite heterogeneous, depending on how the assay is constructed and how the results are interpreted (Table 1).

A variety of technical issues may contribute to how well a marker correlates with biologic and clinical end points. These issues must be described and understood for each specific marker and for each specific use ("use" is described below). For example, mutations in p53 may be detected by sequence analysis of DNA, by single-strand conformational polymorphism screening of DNA, or by immunohistochemical analysis of tissue for p53 protein (the latter is detectable only if mutations have altered the protein to protect it from rapid degradation) (13). Each of these assays may produce different results and conclusions regarding the clinical utility of presumed p53 mutations as a prognostic factor.

Alternatively, one reagent may be used in different assay formats. For example, a monoclonal antibody may be used for immunohistochemical studies to detect and semi-quantify tissue

Table 1. Molecules or substances to be assayed, potential assays, and alterations detected for tumor markers

| What is molecule or process that is assayed? | What alteration is assay detecting? | What is assay format?* | What is reagent? | What are conditions? | What is considered a positive signal? |
|--|--|---|--|--|---|
| Gene | Amplification, deletion, mutation, etc. | Southern, CDGE, SSCPE, PCR/sequence, etc. | Probe (full length, partial, primer sequence, etc.) | Stringency, etc. | Depends on test, multiple possibilities |
| RNA | Overexpression, mutation, etc. | Northern, reverse PCR; in situ hybridization, etc. | Same as above | Same as above | Same as above |
| Product (protein, carbohydrate, lipid, etc.) | Overexpression, abnormal glycosylation, abnormal cellular location, etc. | ELISA, EIA, RIA, IRMA, immunohistochemical (immunoperoxidase, fluorescence), etc. | Polyclonal antibody, monoclonal antibody, ligand, etc. | Concentration of reagent, direct versus indirect, etc. | Same as above |
| Process (blood vessel growth, cellular response, etc.) | Presence of new vessels or tissue, increased cellular response, etc. | Immunopathology, in vitro cellular assay, etc. | Multiple possibilities | Same as above | Same as above |

*CDGE = continuous denaturation gel electrophoresis; SSCPE = single-strand conformational polymorphism electrophoresis; PCR = polymerase chain reaction; ELISA = enzyme-linked immunosorbent assay; EIA = enzyme-linked immunoassay; RIA = radioimmunoassay; IRMA = immuno-radiomimetic assay.

antigen expression, or it might be incorporated into an enzyme-linked immunoassay to provide a quantitative measure of the same antigen in tissue suspension, or a solution (14). These two techniques will produce different analyses of the same marker, with potentially variable clinical significance. Thus, each claim should be based on independent studies that demonstrate the utility of that marker in the manner in which it was tested, rather than on assumptions that one method provides the same correlation with end points and outcomes as another.

Reagents, Conditions, and Signal Detection Systems Used in Assay Formats

Detection and quantification of a marker with different reagents may not produce identical results, even if these reagents are used in similar assay formats (Table 1) (15). For example, in one study, immunohistochemical staining was performed on consecutive tissue sections from the same blocks, using three different monoclonal antibodies directed against separate epitopes on the same breast cancer-associated mucin-like antigen (MUC-1) (16). In this study, associations between immunohistochemical positivity and clinical outcomes were distinct for each of the three antibodies, presumably because expression of each epitope differed from the other two (16).

A specific assay for an individual marker, even if performed in a uniform fashion, may be interpreted differently with the use of various systems of signal analysis. For example, several methods for scoring immunohistochemical staining have been proposed. The same slide may be read by light microscopy with human interpretation or by computer-based scanning. One might consider only nuclear staining critical and ignore all other pat-

terns (e.g., cytosolic or membrane) or vice versa. Some scoring systems consider only whether cells are positive or negative. In these systems, results are expressed as "percent positive cells." In other scoring systems, intensity of staining is factored in with the percentage of cells positive to create an immunohistochemical score or index (17). Comparison of results from these two separate analyses might lead to quite different conclusions.

Changes from normality may be expressed in a continuous or categorical fashion. Many different methods to distinguish an abnormal state from a normal or previous base-line condition have been proposed, and these methods may be assay specific (3,4,18-20). How a cutoff value is chosen and which cutoff value is used may substantially alter the association with clinical outcomes.

Specimen Source on Which Assay Is Performed

One assay may be used to detect a marker in different specimen types. Specimen types and methods of collection that might influence assay results are listed in Table 2. For example, an enzyme-linked immunoassay for the HER-2/neu proto-oncogene product may be used to detect the antigen in fresh or frozen tissue or to detect a portion of that same antigen in plasma (14). The biologic and clinical significance of a marker detected in association with a cell (e.g., in the cytosol or membrane) may be very different from that of the same marker when it is detected as a soluble factor in fluid. Furthermore, results obtained from a cellular needle aspirate might be different from those obtained using the same assay on the identical cells collected in a large biopsy specimen in which tissue architecture is preserved.

Different strategies of specimen preparation and storage may radically alter assay results. Examples of common storage stra-

Table 2. Factors that affect specimen source and methods of collection for tumor marker assays

| Type of specimen/materials | Methods of collection | Methods of preparation | Methods of storage |
|--|---|--|---|
| <i>Assays that detect cellular-based marker (marker is in or on cell)</i> | | | |
| Tissue-based cells | Excisional or incisional biopsy Large-bore needle biopsy | Fresh, frozen, or fixed. If fixed, what was fixative (e.g., formalin, Zenker's, etc.)? | Intact block (frozen or fixed) After microsectioning and placed on glass slide After cellular disruption as pellet, powder, or individual molecular preparation (e.g., DNA) Room temperature (20 °C) Refrigerated (4 °C) Frozen (-20 °C, -70 °C, liquid nitrogen) Time of storage |
| Suspended cells | Fine-needle aspiration Scraping of skin or mucosa Collection of normal circulating cells in blood Collection of exfoliated cells in secreted contents (sputum, urine, stool, nipple aspirate, or milk) or in body fluids (blood, cerebrospinal fluid, effusions) | Same as above | Same as above |
| <i>Assays that detect noncellular-based marker (marker is soluble or is in cell-free suspension)</i> | | | |
| Circulating fluid | Serum Plasma | Method of anticoagulation (EDTA, heparin, etc.) | Room temperature (20 °C) Refrigerated (4 °C) Frozen (-20 °C, -70 °C, liquid nitrogen) Has sample been frozen/thawed? If so, number of cycles Time of storage |
| Secreted or body fluids | Sputum, urine, stool, nipple aspirate, or milk effusions | Method of anticoagulation, if any (EDTA, heparin, etc.), concentration, lyophilization, etc. | Same as above |

gies are listed in Table 2. Many tumor marker assays, especially those that are based on antibody-antigen reactions, are much more sensitive in fresh or frozen specimens. Moreover, in fixed tissues, antigen "survival" may vary with different fixatives. Many assays are affected not only by the types of fixatives or anticoagulants used in preparing the specimen, but also by conditions of storage (different temperatures, exposure to air, etc.) and the length (time) of storage (months versus years) (21). In addition, for many soluble markers, multiple freezing and thawing cycles may decrease reactivity, depending on the specific assay. Furthermore, even if collection and storage methods are standardized, sampling differences (e.g., reviewing multiple tissue section slides as opposed to one per case) may substantially alter results.

Disease for Which Marker Is Being Assayed

One tumor marker, even if defined precisely and assayed uniformly in one disease, may not have any or may have a completely different clinical utility in another cancer (*see below*, "Clinical Uses"). Although it is important to designate the disease for which the marker is being evaluated, the TMUGS should be applicable, with minor modifications, to most cancers.

Evaluating Clinical Utility of Specific Tumor Markers

Overview of the TMUGS

The TMUGS is designed to determine whether the weight of available evidence suggests that knowledge of marker data for an individual patient can be reliably used to make clinical decisions that will improve outcome. A secondary use of the TMUGS would be to design an efficient clinical study of a new or an established marker. Tumor marker data might be useful in at least nine separate clinical situations, designated "uses" for the purpose of the TMUGS (*see under "Utility" in column 1, Fig. 1*). For each potential use, tumor marker data should be evaluated to determine whether they are reliably associated with the biologic process being considered and whether that association predicts a future end point regardless of whether that knowledge has any clinical therapeutic relevance.

Ultimately, knowledge of tumor marker data should lead to a clinical decision that results in a more favorable clinical outcome than if the marker results were not known. The TMUGS is designed to determine the utility of the marker in helping to reliably make clinical decisions that result in improvements in one of four clinical outcomes: overall survival, disease-free survival, quality of life, or cost of care (22).

To estimate a marker's utility for any of the nine uses, separate, semiquantitative utility scales were developed for two categories: 1) biologic process and end points and 2) clinical outcomes. The assigned utility score in these scales reflects the reviewer's summary of the state of the art for that marker in each situation. To justify the assignment of a specific utility score, the reviewer provides an estimate of the level of evidence to support his/her evaluation with an associated list of references. The reviewer might be evaluating only a single study of a

given marker, or he/she might use the TMUGS to perform a comprehensive overview of available studies. In the following sections, each of these components of the TMUGS is described in greater detail.

Clinical Uses

A tumor marker may be useful in evaluation and treatment decisions in one or more clinical situations, including risk assessment, screening, differential diagnosis, prognosis, and monitoring disease course (Fig. 1). Application of screening and/or preventive strategies is most efficient in populations of patients who are at highest risk of developing the anticipated event (23). The recent identification of several "cancer susceptibility genes" will provide the opportunity to estimate cancer risk more precisely (24). Screening for signs of malignancy before it becomes clinically manifest is valuable if early treatment of the cancer in question substantially reduces morbidity and mortality.

Histopathologic diagnosis represents the paradigm of a tumor marker. If histopathology is equivocal, tumor markers might help to distinguish malignant from benign tissue after biopsy, to distinguish hematologic from epithelial from mesenchymal cancers, and potentially even to distinguish one tissue type from another (25).

Prognosis for patients with established primary or metastatic cancer is defined as the prediction of future behavior of an established malignancy, either in the absence of or after application of therapy (local or systemic). We have followed the suggestion of McGuire and Clark (26), who proposed that prognostic factors be divided into two categories: 1) those that predict relapse or progression independent of future treatment effects (which we will designate "prognostic" factors) and 2) those that predict response or resistance to a specific therapy (which we will designate "predictive" factors).

A factor can be both prognostic (of likelihood of relapse and/or progression) and predictive (of likelihood of benefit from therapy). For example, untreated patients with newly diagnosed estrogen receptor (ER)-negative primary breast cancer have a higher risk of relapse over a shorter period of time than do ER-positive patients with a similar disease stage, presumably because ER is associated with either metastatic and/or growth potential (27,28). In this case, ER is "prognostic." On the other hand, the antiestrogen tamoxifen is more effective in preventing breast cancer recurrences in ER-positive patients than in ER-negative patients (29). In this case, ER is "predictive" of benefit from tamoxifen.

During treatment and subsequent follow-up, marker results might be used to monitor patients. One might monitor those patients who have no (detectable) evidence of disease to detect impending relapse (Fig. 1, *see "Detect relapse in patients with no evidence of disease after therapy for primary or recurrent disease"*) or those with detectable disease to determine whether the current therapeutic regimen is effective (Fig. 1, *see "Follow detectable disease"*).

Marker Correlation With Biologic Processes and Biologic End Points and Marker Use Leading to More Favorable Clinical Outcomes

Ultimately, a tumor marker result is clinically useful only if knowledge of its status promotes a change in practice that

favorably affects a clinical outcome. Therefore, tumor marker clinical utility cannot be separated from the known benefits of therapy for that condition. However, research regarding tumor markers, even for diseases and uses for which no effective therapy is currently available, should continue, with the understanding that tumor marker data will be valuable in the context of future therapeutic advances. In the TMUGS, we have separated the tumor marker assay's association with a biologic process or end point from its clinical utility (Fig. 1, see "Utility"). To avoid confusion, we have designated the association between marker results and disease behavior (e.g., development of cancer in the "Risk" category or progression of cancer in the "Prognosis" categories) as a "Correlation With Biologic End Point." In contrast, we have designated use of marker results to dictate a clinical decision that is beneficial to the patient as a "Favorable Clinical Outcome." Thus, we use the term "outcomes" only if it specifically refers to a therapeutic decision that leads to a superior clinical result (22).

For any diagnostic test, it is important to know how well the test performs in populations with and without the disease in question (30,31). These performance characteristics include the sensitivity and specificity and the positive and negative predictive values of the assay (32). Performance characteristics can be used to objectively determine the associations of assay results with biologic processes and end points and, consequently, the effects in producing more favorable clinical outcomes.

Marker association with biologic process and biologic end point. For a marker to be of value in a clinical setting, it must reflect the biologic process with which it is putatively associated. However, association with a biologic process, even if quite good, does not necessarily imply clinical utility. If the knowledge of the marker result does not lead to a decision in clinical practice that results in a more favorable clinical outcome (overall survival, disease-free survival, quality of life, and cost of care), then its use in routine clinical practice is discouraged (22). Nonetheless, on an investigational basis, correlations with biologic end points are still important to ascertain for the marker in various uses, since these correlations might be extremely valuable if and when therapeutic advances for the disease are made. Subsequent and separate evaluations can then be performed with regard to whether tumor marker data have utility in making clinical decisions that result in a favorable clinical outcome.

Two features must be considered for this category: 1) Do the assay results reliably reflect the biologic process or change for which the assay is developed? 2) Do the assay results predict the biologic end point under consideration? For example, immunohistochemical staining for the p53 protein in fixed carcinoma tissue correlates reasonably well with mutations in the p53 gene (33). In this case, the assay for one marker, immunohistochemical staining for increased levels of p53 protein, reflects the gold standard assay, sequencing DNA for base-pair mutations. Nonetheless, it is possible that such antibodies cross-react with other molecules or do not react with p53 at all. Thus, before any clinical utility of staining with these assays can be determined, it is important to demonstrate that they recognize mutated p53 protein with reasonable sensitivity and specificity (13).

However, knowledge that an assay reflects a biologic process does not mean that results of that assay predict future behavior of the tumor. For example, it is now well established that up to 50% of breast cancers exhibit evidence of p53 mutations. However, the data from studies that attempt to determine the prognostic value of mutated p53 for a higher likelihood of recurrence and mortality are mixed (13).

In summary, the category describing association with biologic processes and biologic end points is intentionally included as the first "utility" to be evaluated for each tumor marker in each use. Association of tumor marker results with a biologic process may help set an investigational agenda by providing insight for future marker studies that address biologic end points. Moreover, such studies should also contribute to development of more effective therapies for cancers in which tumor markers are associated with biologic end points.

Marker leading to decision in clinical practice that results in favorable outcome. Although association with a biologic process is of interest for investigational purposes, standard use of a marker in routine clinical practice should be recommended only if the marker reliably adds to the clinician's judgment during clinical decision-making, resulting in a more favorable clinical outcome for the patient. These favorable outcomes are increased overall survival, increased disease-free survival, improved quality of life, and/or reduced cost of care (Fig. 1) (22).

Overall survival is defined as the overall time a patient will live. By virtue of earlier detection, tumor marker data may artificially increase perceived survival after diagnosis, resulting in lead-time bias. However, true overall survival may be increased because earlier treatment might lead to better chances of living longer. Thus, measurement of overall survival must be performed properly, such as in a prospective, randomized trial. Nonetheless, prolongation of overall survival is relatively straightforward to evaluate and is universally accepted as a desired end point. In contrast, prolongation of disease-free survival commonly serves as an assumed surrogate for either prolonged overall survival or improved quality of life. However, depending on the disease and use in question, disease-free survival may not necessarily be a reliable indicator of either. Quality of life is a more meaningful end point than disease-free survival, but it is often more difficult to quantify objectively (34-36). Reduction in cost of care is an equally important goal of using tumor markers. Reduced costs of care may occur by replacement of a more expensive test with an equally reliable but more economic tumor marker result.

Utility Scales to Evaluate Tumor Markers

As a critical component of the TMUGS, we developed semi-quantitative utility scales that describe the reviewers' interpretation of the current status of a marker for biologic and clinical outcomes for each use (Tables 3, 4). In general, the values in these scales reflect the performance characteristics for the assay in relationship to the respective use and the biologic process and biologic end point and/or favorable clinical outcome.

Evaluation of the marker using the utility scale requires a review of available data and assignment of a utility score for the marker in each specific use. Assignment of a utility score to a marker for a given use is not irrevocable, since this process re-

Table 3. Utility scale to evaluate tumor marker for correlation with biologic process and biologic end points

| Utility scale | Explanation of scale |
|---------------|--|
| 0 | Marker does not correlate with process, or marker does not correlate with expected end point. |
| NA | Data are not available regarding marker correlation with process or end point for that use. |
| +/- | Preliminary data are suggestive that assay correlates with process, or with end point, but substantially more definitive studies are required. |
| + | Assay probably associated with process or end point, but additional confirmatory studies are required. |
| ++ | Definitive studies demonstrate that assay reflects process or end point. |

quires subjective review of available data (*see below*, "Levels of Evidence. . .") and since new data regarding either the marker or the therapy may (and, indeed should) change the utility.

We have developed two separate scales to reflect the distinction between evaluation of a marker's performance characteristics in regard to biologic process and biologic end points (Table 3) and the utility of marker data to produce more favorable clinical outcomes (Table 4). Although these scales are similar, they have different functions. As described, association of marker with biologic process and biologic end points does not imply clinical utility, although it may. Thus, utility scores from "0" to "++" are assigned for the columns describing biologic process and biologic end points (Table 3).

However, a critical evaluation of the marker in the context of existing therapeutic benefits must be made to meet the criteria for a marker to be considered "standard practice." Thus, the scale to evaluate markers for clinical outcomes includes an additional score, "+++" (Table 4). In the favorable clinical outcomes scale (Table 4), only scores "++" and "+++" are sufficient for a marker to be considered standard clinical practice. In contrast, assignment of a "0" to a marker in a given use implies that sufficient data are available to document that, even if the marker had appeared promising in preliminary studies and/or even if it correlates highly with a biologic end point, it has no utility for that use when critically investigated in the context of available therapeutic options. A "0" may be assigned because the marker

data are so poor that they do not correlate with the biologic process or biologic end point to be useful (e.g., they may have been assigned a "+/-" or "+" in these categories). Alternatively, in the favorable clinical outcomes columns, "0" might be assigned because the therapeutic options available for that disease are insufficient to render tumor marker data of any utility, even if the marker is associated with the biologic process or biologic end point.

One example of the former situation is that of tissue carcinoembryonic antigen (CEA) expression and breast cancer. Immunohistochemical staining reliably reflects the expression of CEA. Therefore, we would assign a utility score of "++" to the biologic process correlation column in the "Prognosis: Primary" row in Fig. 1. However, CEA expression is only weakly predictive, if at all, of a higher risk of recurrence in patients with stage I or II breast cancer (37). Thus, even though adjuvant systemic therapy does reduce the odds of recurrence for patients with breast cancer, the power to distinguish favorable from unfavorable prognosis with CEA staining is so weak that this technique has no role in clinical care. Therefore, we would assign a "0" in the column for correlation with biologic end point and in all the succeeding columns (overall survival, disease-free survival, quality of life, and cost of care) as well.

An example of a second reason to assign a "0" is detection of H-ras mutations in non-small-cell lung cancer (NSCLC). These mutations are correlated with worse outcome in patients with

Table 4. Scale to evaluate utility of tumor markers for favorable clinical outcomes

| Utility scale | Explanation of scale |
|---------------|--|
| 0 | Marker has been adequately evaluated for a specific use, and the data definitively demonstrate it has no utility. The marker should not be ordered for that clinical use. |
| NA | Data are not available for the marker for that use because marker has not been studied for that clinical use. |
| +/- | Data are suggestive that the marker may correlate with biologic process and/or biologic end point, and preliminary data suggest that use of the marker may contribute to favorable clinical outcome, but more definitive studies are required. Thus, the marker is still considered highly investigational and should not be used for standard clinical practice. |
| + | Sufficient data are available to demonstrate that the marker correlates with the biologic process and/or biologic end point related to the use and that the marker results might affect favorable clinical outcome for that use. However, the marker is still considered investigational and should not be used for standard clinical practice, for one of three reasons: <ol style="list-style-type: none"> 1) The marker correlates with another marker or test that has been established to have clinical utility, but the new marker has not been shown to clearly provide any advantage. 2) The marker may contribute independent information, but it is unclear whether that information provides clinical utility because treatment options have not been shown to change outcome. 3) Preliminary data for the marker are quite encouraging, but the level of evidence (<i>see below</i>) is lacking to document clinical utility. |
| ++ | Marker supplies information not otherwise available from other measures that is helpful to the clinician in decision-making for that use, but the marker cannot be used as sole criterion for decision-making. Thus, marker has clinical utility for that use, and it should be considered standard practice in selected situations. |
| +++ | Marker can be used as the sole criterion for clinical decision-making in that use. Thus, marker has clinical utility for that use, and it should be considered standard practice. |

newly diagnosed stage II NSCLC (38). Thus, one would assign either a “+” or “++” for these markers in the columns for “Marker association with biologic process and biologic end point” in the row for “Prognosis: Primary.” However, results from several clinical trials have failed to demonstrate definitively that overall survival, disease-free survival, quality of life, or cost of care for patients with newly diagnosed stage I or II NSCLC is improved with application of currently available adjuvant systemic therapies after initial surgery (39,40). Thus, these markers would be assigned a “0” for this use for these outcomes at that time, even though they would be assigned “+” or “++” in the preceding columns for correlation with biologic process and biologic end points. If effective adjuvant therapies are developed, then these markers might be reassigned more favorable scores.

If the marker has not been studied or data are insufficient to make a judgment for that clinical use, the marker is assigned a utility score of “NA.” This category is for markers for which theoretical bases exist to hypothesize a potential utility for a given use, but for which the appropriate studies have not been performed. For example, it is anticipated that assays to detect abnormalities in the recently cloned breast cancer-associated gene BRCA1 will soon be available (41,42). Theoretically, these assays should identify subjects who are very likely to develop breast cancer. Although it is not clear how sensitive or specific these assays will be, it is likely that they will be assigned a score of “+” or even “++” for the biologic process column. However, it is suspected (and indeed it is likely) that not all abnormalities in this gene will be associated with the development of breast cancer (24,43,44). Thus, we would assign these assays a score of “NA” for the biologic end point column until properly designed clinical studies are performed.

More importantly, it is not known whether currently existing preventive strategies, such as surgical organ ablation or chemoprevention, will reduce the odds of developing cancer in affected individuals (24,43,44). Thus, again, we would assign utility scores of NA for the clinical outcomes column until appropriate studies are performed.

The categories of “+/-” and “+” are the next steps after “NA” on a continuum of investigations to determine the utility of a marker in a given use. This designation would be assigned during early evaluation of a marker, when results from only a few preliminary studies are available or when results from several studies conflict. The “+” category implies that the marker is promising but cannot be considered standard clinical practice. Three possible scenarios exist for assignment of a “+” score rather than a more definitive “++” for biologic correlations or “++” or “+++” for favorable clinical outcomes.

1) The marker correlates with another marker or test that has been established to have clinical utility, but the new marker has not been shown to clearly provide any advantage. For example, several breast cancer tissue-based markers (cathepsin D, tissue neovascularization) appear to correlate with the presence or absence of lymph node involvement and even a high risk of recurrence (1). Therefore, these markers would be assigned “+” or “++” for both the biologic process and biologic end points columns. However, results from the many reported studies are inconsistent with regard to whether any of these markers

provides independent information beyond what is already known from clinical stage and lymph node status. Thus, these markers should not be used to make treatment decisions and would be assigned scores of “+” for any of the clinical outcomes columns. However, it is possible that results of future investigations will suggest that these markers may either complement or even replace lymph node status (utility score = “++” or “+++”, respectively) or that they do not provide any additional information (utility score = 0).

2) The marker may contribute independent information, but it is unclear whether that information provides clinical utility because treatment options have not been shown to change outcome. In this case, the marker may have a “+” or “++” score in the biologic process and biologic end point columns (*see above*) but only a “+” for clinical utility. For example, as noted previously, several tissue-based markers are reliable prognostic factors in patients with newly diagnosed NSCLC (45). Although no clinical trial has established that early systemic adjuvant therapy for NSCLC patients improves any of the four stated clinical outcomes, subgroup analysis has suggested that perhaps there are populations of patients within these trials for whom adjuvant chemotherapy improves disease-free survival (39,40). It is possible that more effective therapies that have been developed subsequent to these randomized trials would result in a favorable outcome for these patients. However, that assumption must be demonstrated in a properly controlled, prospective trial or in large overview analyses of many studies.

3) Preliminary data for the marker are quite encouraging, but the level of evidence (*see below*) is lacking to document clinical utility. In this case, clinical data have been generated, and results are not uniformly positive or negative or are not sufficient to be considered “definitive.” For example, several studies (46,47) have suggested that HER-2/neu overexpression may predict relative sensitivity or resistance to adjuvant chemotherapy. Although intriguing, these results must be interpreted in light of the well-described hazards of retrospective subset analysis, and they require further confirmation in other studies (*see “Levels of Evidence to Assign Utility Score to Marker For Use” below*). Thus, we would assign HER-2/neu expression a score of “+” for the use of “Prognosis: Predict response to therapy: Primary” (Fig. 1).

The final two categories “++” and “+++” imply that the marker has clinical utility. The “++” category is the one into which many clinically useful markers will be placed. In this setting, the marker complements other information (history, physical examination, radiography, routine histopathology, other markers) used by the clinician to judge the patient’s status and to decide which avenue of practice will be most beneficial to the patient.

For example, the performance characteristics of a rising CEA level during follow-up of a patient who has completed primary and adjuvant therapy for either colon or breast cancer are sufficient to reliably predict future recurrence within the following few months to years (46,47). Thus, for either disease, we would assign a score of “++” for the “Correlation with biologic processes and biologic end points” columns in the row designated “Monitor course: Detect relapse in patient with no evidence of disease after therapy for primary or recurrent disease.”

However, the clinical utility of monitoring CEA levels to detect early relapse may be strikingly different for these two diseases. Although the issue is controversial, there is substantial evidence that, in certain colon cancer patients with a rising CEA level, isolated hepatic metastases can be identified and resected, with an apparent cure rate of 20% (46,48-50). We would assign a utility score of “++” for serial monitoring of CEA in such patients, since the marker is used in the context of other diagnostic tests (such as computed tomography scans) to indicate a clinical approach that results in a more favorable outcome, at least in some patients. In contrast, although a rising marker predicts relapse in breast cancer patients, such patients are not apparently better treated because of this knowledge (51). Therefore, serial monitoring of breast cancer patients, in the same situation as that of colon cancer patients, is assigned a “0” or, at best, a “+/-.”

Few markers can be used as the sole criteria for clinical decision-making; therefore, most will be assigned “++” or less. Nonetheless, in some situations, changes in clinical practice are indicated on the basis of marker results alone. Perhaps the best example is the use of α -fetoprotein and β -human chorionic gonadatropin in men with testicular cancer. An elevated or rising circulating level of either or both of these markers after a patient has been rendered free of detectable disease (by surgery or chemotherapy) is pathognomonic for recurrence and is an indication for treatment, regardless of whether disease can be detected by other means (52-54). Markers assigned this score should be considered standard practice in the evaluation and monitoring of all patients with the disease in question.

Levels of Evidence to Assign Utility Score to Marker for Use

Initially, the TMUGS was designed to develop practice guidelines for tumor markers (see “Notes” section). Extensive reviews of practice guideline development are published elsewhere (55-58). However, a cornerstone of practice guideline development is a critical review of published investigational data to develop and support the conclusions of the reviewer. Thus, the reviewer can place the available data into one of several “Levels of Evidence.” These levels are categories that

define the quality of data that exist on which the utility score is based.

We have modeled our “Levels of Evidence Scale” on that proposed by the Canadian Task Force on the Periodic Health Examination (59). In that scale, used to develop practice guidelines for specific therapeutic options, levels range from I to V. Level I evidence is considered definitive and is obtained from a single, high-powered, prospective, randomized, controlled trial or from a meta-analysis or overview of multiple, well-designed studies. Level V evidence is considered quite weak and is derived from case reports and clinical examples. Levels II-IV represent various degrees between these two extremes.

The modified levels of evidence scale for tumor markers are provided in Table 5. Study designs of tumor markers often differ from those of new therapeutic agents (19,20,60). Clinical specimens may have been collected from patients in a number of settings. However, as with studies of therapeutic modalities, the design of tumor marker studies will place the obtained results within different levels.

Of note, the statistical analysis of each study may profoundly reflect which of the levels of evidence it is considered to represent. Statistical analysis differs from the technique of interpretation for a given assay, as discussed above in the marker definition sections. Just as many different methods exist to collect tumor marker data, many statistical techniques may also be used to evaluate similar observations, thus resulting in heterogeneous conclusions (19,20,61-65). For example, many markers are frequently reported to be associated with clinical outcome. However, when evaluated in the context of previously reported (and/or accepted) factors in multivariate analysis, these results may not provide additional information. Furthermore, different techniques of multivariate analysis may also vary, and each must be assessed critically (20).

Conclusion

We recognize that the proposed TMUGS is complex. Use of the TMUGS for evaluation of available data to assign a utility scale score requires human judgment to compile and assess the magnitude and clinical importance of the observed benefit

Table 5. Levels of evidence for grading clinical utility of tumor markers

| Level | Type of evidence |
|-------|---|
| I | Evidence from a single, high-powered, prospective, controlled study that is specifically designed to test marker or evidence from meta-analysis and/or overview of level II or III studies. In the former case, the study must be designed so that therapy and follow-up are dictated by protocol. Ideally, the study is a prospective, controlled randomized trial in which diagnostic and/or therapeutic clinical decisions in one arm are determined at least in part on the basis of marker results, and diagnostic and/or therapeutic clinical decisions in the control arm are made independently of marker results. However, study design may also include prospective but not randomized trials with marker data and clinical outcome as primary objective. |
| II | Evidence from study in which marker data are determined in relationship to prospective therapeutic trial that is performed to test therapeutic hypothesis but not specifically designed to test marker utility (i.e., marker study is secondary objective of protocol). However, specimen collection for marker study and statistical analysis are prospectively determined in protocol as secondary objectives. |
| III | Evidence from large but retrospective studies from which variable numbers of samples are available or selected. Therapeutic aspects and follow-up of patient population may or may not have been prospectively dictated. Statistical analysis for tumor marker was not dictated prospectively at time of therapeutic trial design. |
| IV | Evidence from small retrospective studies that do not have prospectively dictated therapy, follow-up, specimen selection, or statistical analysis. Study design may use matched case-controls, etc. |
| V | Evidence from small pilot studies designed to determine or estimate distribution of marker levels in sample population. Study design may include “correlation” with other known or investigational markers of outcome but is not designed to determine clinical utility. |

(hence, our utility scales) and the likelihood that these are not due to play of chance (hence, our Levels of Evidence Scale). Such a process will require interpretation of both the tumor marker data and the therapeutic data (hence, the development of two utility scales).

Laboratory and clinical investigators may also wish to use the TMUGS while considering study designs for new markers. Admittedly, adherence to such a system in order to achieve acceptance for a marker will require commitment to long-range study designs that may be time-consuming and expensive. However, with the proliferation of molecular and immunologic techniques, many putative markers are being proposed and investigated. Indeed, individual users of the system may wish to extract only certain features for their own particular objectives.

Admittedly, the TMUGS as it is presented is imperfect. We hope that this publication generates a public dialogue regarding specific issues, such as what is considered "standard practice." Currently, results from unproven markers are being made available to clinicians, with no guidelines as to the reliability of the assay or to the evidence regarding how or if it should be used to make clinical decisions. The TMUGS may serve as a framework in which tumor markers, like therapeutic agents, can be appropriately tested and introduced into clinical practice in order to improve patient outcomes, protect patient interests, and permit more cost-efficient application of effective therapies.

References

- (1) Hayes DF. Tumor markers for breast cancer. *Ann Oncol* 1993;4:807-19.
- (2) Simon R. Design and conduct of clinical trials. In: DeVita VT Jr, Hellman S, Rosenberg SA, editors. *Cancer: principles and practice of oncology*. 4th ed. Philadelphia: Lippincott, 1993:418-40.
- (3) Lin DY, Fischl MA. Evaluating the role of CD4-lymphocyte counts as surrogate endpoints in human immunodeficiency virus clinical trials [see comment citation in Medline]. *Stat Med* 1993;12:835-42.
- (4) Hillis A, Seigel D. Surrogate endpoints in clinical trials: ophthalmologic disorders. *Stat Med* 1989;8:427-30.
- (5) Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177-82.
- (6) Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989;244:707-12.
- (7) Hayes DF, Cirincione C, Carney W, Rodrigue S, Berry D, Younger J, et al. Elevated circulating HER-2/neu related protein (NRP) is associated with poor survival in patients with metastatic breast cancer [abstract]. *Proc ASCO* 1993;12:58a.
- (8) Leitzel K, Teramoto Y, Sampson E, Mauceri J, Langton BC, Demers L, et al. Elevated soluble c-erbB-2 antigen levels in the serum and effusions of a proportion of breast cancer patients. *J Clin Oncol* 1992;10:1436-43.
- (9) Leitzel K, Teramoto Y, Konrad K, Chinchilli VM, Volas G, Grossberg H, et al. Elevated serum c-erbB-2 antigen levels and decreased response to hormone therapy of breast cancer. *J Clin Oncol* 1995;13:1129-35.
- (10) Disis ML, Calenoff E, McLaughlin G, Murphy AE, Chen W, Groner B, et al. Existent T-cell and antibody immunity to HER-2/neu protein in patients with breast cancer. *Cancer Res* 1994;54:16-20.
- (11) Lupu R, Lippman ME, William L. McGuire Memorial Symposium. The role of erbB2 signal transduction pathways in human breast cancer. *Breast Cancer Res Treat* 1993;27:83-93.
- (12) Frebourg T, Kassel J, Lam KT, Gryka MA, Barbier N, Andersen TI, et al. Germ-line mutations of the p53 tumor suppressor gene in patients with high risk for cancer inactivate the p53 protein. *Proc Natl Acad Sci U S A* 1992;89:6413-7.
- (13) Thor AD, Yandell DW. Prognostic significance of p53 overexpression in node-negative breast carcinoma: preliminary studies support cautious optimism [editorial] [see comment citation in Medline]. *J Natl Cancer Inst* 1993;85:176-7.

- (14) Carney WP, Hamer PJ, Petit D, Retos C, Greene R, Zabrecky JR, et al. Detection and quantitation of the human neu oncogene. *J Tumor Marker Oncol* 1991;6:53-72.
- (15) Taylor CR. An exaltation of experts: concerted efforts in the standardization of immunohistochemistry. *Hum Pathol* 1994;25:2-11.
- (16) Hayes DF, Mesa-Tejada R, Papsidero L, Croghan G, Korzun A, Norton L, et al. Prediction of prognosis in primary breast cancer by detection of a high molecular weight mucin-like antigen using monoclonal antibodies DF3, F36/22, and CU18: a Cancer and Leukemia Group B study. *J Clin Oncol* 1991;9:1113-23.
- (17) McClelland RA, Berger U, Miller LS, Powles TJ, Coombes RC. Immunocytochemical assay for estrogen receptor in patients with breast cancer: relationship to a biochemical assay and to outcome of therapy. *J Clin Oncol* 1986;4:1171-6.
- (18) Clark GM, Dressler LG, Owens MA, Pounds G, Oldaker T, McGuire WL. Prediction of relapse or survival in patients with node-negative breast cancer by DNA flow cytometry. *N Engl J Med* 1989;320:627-33.
- (19) Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology [editorial]. *Br J Cancer* 1994;69:979-85.
- (20) George SL. Statistical considerations and modeling of clinical utility of tumor markers. In: Hayes DF, editor. *Hematology/Oncology Clinics of North America: tumor markers in adult solid malignancies*. Philadelphia: Saunders, 1994:457-70.
- (21) Bromley CM, Palechek PL, Benda JA. Preservation of estrogen receptor in paraffin sections. *J Histotechnol* 1994;17:115-8.
- (22) American Society of Clinical Oncology Outcomes Working Group. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 1996;14:671-9.
- (23) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually [see comment citations in Medline]. *J Natl Cancer Inst* 1989;81:1879-86.
- (24) Garber JE. Markers of risk for human malignancies. In: Hayes DF, editor. *Hematology/oncology clinics of North America: tumor markers in adult solid malignancies*. Philadelphia: Saunders, 1994:471-83.
- (25) Hayes DF. What are tumor markers, and when should they be used? In: American Society of Clinical Oncology. *Educational book: 30th annual meeting*. Dallas: American Society of Clinical Oncology, 1994:138-49.
- (26) McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary-node-negative breast cancer [see comment citations in Medline]. *N Engl J Med* 1992;326:1756-61.
- (27) Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors [prior annotation incorrect]. *N Engl J Med* 1989;320:479-84.
- (28) Fisher B, Redmond C, Dimitrov NV, Bowman D, Legault-Poisson S, Wickerham DL, et al. A randomized clinical trial evaluating sequential methotrexate and fluorouracil in the treatment of patients with node-negative breast cancer who have estrogen-receptor-negative tumors [prior annotation incorrect]. *N Engl J Med* 1989;320:473-8.
- (29) Early Breast Cancer Trialists' Collaborative Group T. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet* 1992;339:1-15, 71-85.
- (30) Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-98.
- (31) Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Diag Radiol* 1982;143:29-36.
- (32) Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH, editors. *Biostatistics in clinical medicine*. New York: Macmillan, 1987.
- (33) Thor AD, Moore DH 2d, Edgerton SM, Kawasaki ES, Reihnsaus E, Lynch HT, et al. Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. *J Natl Cancer Inst* 1992;84:845-55.
- (34) Gelber RD, Goldhirsch A, Humy C, Bernhard J, Simes RJ. Quality of life in clinical trials of adjuvant therapies. International Breast Cancer Study Group (formerly Ludwig Group). *Monogr Natl Cancer Inst* 1992;11:127-35.
- (35) Donovan K, Sanson-Fisher RW, Redman S. Measuring quality of life in cancer patients. *J Clin Oncol* 1989;7:959-68.
- (36) Porzolt F, Tannock I. Goals of palliative cancer therapy. *J Clin Oncol* 1993;11:378-81.
- (37) Robertson JF, Ellis IO, Bell J, Todd JH, Robins A, Elston CW, et al. Carcinoembryonic antigen immunocytochemistry in primary breast cancer. *Cancer* 1989;64:1638-45.
- (38) Slebos RJ, Evers SG, Wagenaar SS, Rodenhuis S. Cellular protooncogenes are frequently amplified in untreated non-small cell lung cancer. *Br J Cancer* 1989;59:76-80.

- (39) Sandler AB, Buzaid AC. Lung cancer: a review of current therapeutic modalities. *Lung* 1992;170:249-65.
- (40) Jaklitsch MT, Strauss GM, Sugarbaker DJ. Neoadjuvant and adjuvant therapy in the management of locally advanced non-small-cell lung cancer. *World J Surg* 1993;17:729-34.
- (41) Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 1994;266:120-2.
- (42) Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66-71.
- (43) National Advisory Council for Human Genome Research. Statement on use of DNA testing for presymptomatic identification of cancer risk. *JAMA* 1994;271:785.
- (44) Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, et al. A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene: implications for presymptomatic testing and screening. *JAMA* 1995;273:535-41.
- (45) Strauss GM, Skarin AT. Use of tumor markers in lung cancer. In: Hayes DF, editor. *Hematology/Oncology Clinics of North America: tumor markers in adult solid malignancies*. Philadelphia: Saunders, 1994:507-32.
- (46) Kemeny M, Sugarbaker P, Smith T, Edwards B, Shawker T, Vermess M, et al. A prospective analysis of laboratory tests and imaging to detect hepatic lesions. *Ann Surg* 1982;195:163-7.
- (47) Tondini C, Hayes DF, Kufe D. Circulating tumor markers in breast cancer. In: Henderson IC, editor. *Diagnosis and therapy of breast cancer*. Philadelphia: Saunders, 1989:653-74.
- (48) Steele G Jr, Ellenberg S, Ramming K, O'Connell M, Moertel C, Lessner H, et al. CEA monitoring among patients in multi-institutional adjuvant G.I. therapy protocols. *Ann Surg* 1982;196:162-9.
- (49) Minton JP, Hoehn JL, Gerber DM, Horsley JS, Connolly DP, Salwan F, et al. Results of a 400-patient carcinoembryonic antigen second-look colorectal cancer study. *Cancer* 1985;55:1284-90.
- (50) Martin EW Jr, Cooperman M, King G, Rinker L, Carey LC, Minton JP. A retrospective and prospective study of serial CEA determinations in the early detection of recurrent colon cancer. *Am J Surg* 1979;137:167-9.
- (51) Hayes DF, Kaplan W. Evaluation of patients after primary therapy. In: Harris J, Lippman M, Morrow M, Hellman S, editors. *Diseases of the breast*. Philadelphia: Lippincott, 1996:627-45.
- (52) Bartlett NL, Freiha FS, Torti FM. Serum markers in germ cell neoplasms. *Hematol Oncol Clin North Am* 1991;5:1245-60.
- (53) Bosl GJ, Head MD. Serum tumor marker half-life during chemotherapy in patients with germ cell tumors. *Int J Biol Markers* 1994;9:25-8.
- (54) Klein EA. Tumor markers in testis cancer. *Urol Clin North Am* 1993;20:67-73.
- (55) Woolf SH. Practice guidelines: a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med* 1992;152:946-52.
- (56) Tunis SR, Hayward RSA, Wilson MC, Rubin HR, Bass EB, Johnston M, et al. Internists' attitudes about clinical practice guidelines [see comment citations in Medline]. *Ann Intern Med* 1994;120:956-63.
- (57) Kaluzny AD, Rimer B, Harris R. The National Cancer Institute and guideline development: lessons from the breast cancer screening controversy. *J Natl Cancer Inst* 1994;86:901-3.
- (58) Browman GP, Levine MN, Mohide EA, Hayward RS, Pritchard KI, Gafni A, et al. The practice guidelines development cycle: a conceptual tool for practice guidelines development and implementation. *J Clin Oncol* 1995;13:502-12.
- (59) Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979;121:1193-254.
- (60) Levine MN, Browman GP, Gent M, Roberts R, Goodyear M. When is a prognostic factor useful? A guide for the perplexed. *J Clin Oncol* 1991;9:348-56.
- (61) Burke HB, Henson DE. The American Joint Committee on Cancer. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer* 1993;72:3131-5.
- (62) Burke HB. Artificial neural networks for cancer research: outcome prediction. *Semin Surg Oncol* 1994;10:73-9.
- (63) Ravdin PM, Clark GM, Hilsenbeck SG, Owens MA, Vendely P, Pandian MR, et al. A demonstration that breast cancer recurrence can be predicted by neural network analysis. *Breast Cancer Res Treat* 1992;21:47-53.
- (64) Ravdin PM, Clark GM, Hough JJ, Owens MA, McGuire WL. Neural Network Analysis of DNA flow cytometry histograms. *Cytometry* 1993;14:74-80.
- (65) Clark GM, Wenger CR, Beardslee S, Owens MA, Pounds G, Oldaker T, et al. How to integrate steroid hormone receptor, flow cytometric, and other prognostic information in regard to primary breast cancer. *Cancer* 1993;71(6 Suppl):2157-62.

Notes

Supported in part by Public Health Service grant CA64057 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

This manuscript was generated during discussions by an expert panel convened by the American Society of Clinical Oncology (ASCO) to determine clinical practice guidelines for the use of tumor markers. These guidelines will be published elsewhere. We gratefully acknowledge the support of ASCO and its staff during this process. However, this manuscript does not reflect an officially sanctioned policy of ASCO, and each author has elected to contribute independently of his or her participation on the expert panel.

Manuscript received December 21, 1995; revised June 7, 1996; accepted July 15, 1996.