CLINICAL PHARMACOLOGY & THERAPEUTICS
VOLUME 69 NUMBER 3 MARCH 2001

COMMENTARY

Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework

Biomarkers Definitions Working Group *Bethesda, Md*

Technologies used in sequencing of the human genome are dramatically reshaping the research and development pathways for drugs, vaccines, and diagnostics. The growth in the number of molecular entities entering the drug development pipeline has accelerated as a consequence of powerful discovery and screening technologies such as combinatorial chemistry, mass spectrometry, high throughput screening, cell- and tissue-based DNA microarrays, and proteomic approaches.[1] As a consequence, there is an escalating number of therapeutic candidates, which has caused the need for new technologies and strategies to streamline the process to make safe and effective therapies available to patients.

One approach to the achievement of more expeditious and informative therapeutic research is the use of precise clinical measurement tools to determine disease progression and the effects of interventions (drugs, surgery, and vaccines). For example, gene-based approaches such as single nucleotide polymorphism maps are now being developed to distinguish the molecular and cellular basis for variations in clinical response to therapy.[2] Another approach is the use of a wide array of analytical tools to assess biological parameters, which are referred to as biomarkers. Biomarker measurements can help explain empirical results of clinical trials by relating the effects of interventions on molecular and cellular pathways to clinical responses. In doing so, biomarkers provide an avenue for researchers to gain a mechanistic understanding of the differences in clinical response that may be influenced by uncontrolled variables (for example, drug metabolism).

There are a variety of ways that biomarker measurements can aid in the development and evaluation of

Group members and affiliations are listed at the end of the article.
Portions of this article were previously published as a conference proceedings report: Downing GJ, editor. Biomarkers and surrogate endpoints: clinical research and applications. Amsterdam: Elsevier Scientific; 2000. p. 1-7.
Clin Pharmacol Ther 2001;69:89-95.
Received for publication Nov 14, 2000; accepted Dec 31, 2000.
Reprint requests: Gregory J. Downing, DO, PhD, Office of Science Policy, NIH, Bldg 1, Rm 218, One Center Drive, Bethesda, MD 20892.
E-mail: *downingg@od.nih.gov*
13/1/113989
doi:10.1067/mcp.2001.113989

novel therapies.[3] In the initial investigations of therapeutic candidates in humans, biomarkers can provide a basis for the selection of lead compounds for phase 3 clinical trials.[3-5] Biomarkers contribute knowledge about clinical pharmacology and provide a basis for the designing of clinical trials that expeditiously and definitively evaluate safety and efficacy. Biomarkers provide information for guidance in dosing and minimize interindividual variation in response. For example, rapid clearance of ^{99m}Tc-sestamibi, a substrate for P-glycoprotein that is associated with multidrug resistance, has been shown to predict lack of tumor response to adjuvant chemotherapy in some forms of breast cancer.[6] Biomarkers that represent highly sensitive and specific indicators of disease pathways have been used as substitutes for outcomes in clinical trials when evidence indicates that they predict clinical risk or benefit.

Assessment of benefit and risk must be the goal of the development plan for all therapeutic interventions. The most reliable way to assess the clinical impact of a therapeutic intervention (eg, drug, device, surgery, vaccine, biologic agent, and behavioral modality) is through its effect on a well-defined clinical endpoint such as survival, myocardial infarction, stroke, bone fracture, or recurrence of cancer. However, this standard may be impractical for the evaluation of some long-term disease therapies because long periods are required for these clinical endpoints to be achieved and trials with large numbers of patients are needed for their evaluation. Biomarkers that can be reliable substitutes for clinical responses have the potential to improve the efficiency of clinical trials in which long-term disease interventions are evaluated. Biomarkers can be substituted reliably for clinical responses. In other cases, reliable biomarkers have been used as substitutes for clinical endpoints in decision-making situations when a devastating clinical outcome, such as death, represents an ethical dilemma. In recent years, this point has been illustrated in pharmacokinetic studies in which biomarkers such as human immunodeficiency virus (HIV) plasma viral load and CD4 cell counts were used as substitutes for clinical outcomes (for example, death and occurrence of opportunistic infections) in the evaluation of antiviral agents in patients with HIV infection.[7-9]

Accompanying the increased knowledge about biomarkers is an increased appeal of the use of biomarkers as substitutes for clinical outcomes in other diseases. This interest is countered by concerns about the inherent limitations of biomarkers that have been shown by some dramatic failures.[10,11] Among the most notable of these failures is the demonstration by the Cardiac Arrhythmia Suppression Trial (CAST) that suppression of ventricular arrhythmias is not a valid substitute for sudden death after myocardial infarctions.[12] This showed that considerable skepticism about conventional wisdom should accompany the adoption of biomarkers as a substitute for outcomes as the basis for approval of a novel therapy. When biomarkers are intended to be the basis for provisional evaluation and regulatory approval of a drug, they also must be a component of a predetermined strategy that recognizes inability of biomarkers to serve as final proof of clinical efficacy or long-term safety. One strategy for the requirement of a systematic approach with phase 4 trials has recently been proposed for cardiovascular therapies.[13] The reliance on biomarkers as substitutes for outcomes in clinical trials is best justified when adequate safety data are collected and when the results are viewed clearly as a basis for a provisional evaluation that ultimately should be superceded by evidence for clinical benefit.

Despite the shortcomings, there has been long-standing clinical and regulatory acceptance of certain biomarkers as substitutes for clinical endpoints as a basis for drug approval. Reduction of elevated arterial blood pressure has been used for decades to reflect the reduction in the incidence of stroke, congestive heart failure, and subsets of cardiovascular death by antihypertensive drugs.[14] Serum cholesterol level has served as a biomarker for the evaluation of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors that are used to diminish the risk of coronary artery disease.[15] Even in this case, however, new evidence suggests that C-reactive protein, a marker of inflammatory disease activity, may have independent prognostic significance that equals or is additive to that of serum cholesterol levels.[16] Now it appears that some of the protective efficacy of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors could result from the anti-inflammatory effects of these drugs.[17] In some cases, reliance on biomarkers that are prematurely accepted as substitutes for clinical outcome may completely fail to show true benefits of a candidate therapy. For example, in an investigation of interferon gamma on recurrent infections in patients with chronic granulomatous disease, an indicator of phagocytic function failed to show therapeutic effects.[18] However, clear clinical benefit from interferon gamma therapy was achieved as indicated by a substantial reduction in serious infections.

With the anticipated growth in the use of biomarkers in clinical trials, discussions and policies that focus on the use of biomarkers can be best served by a consensus in the use of terminology.[19] Many terms are used to describe measurements of disease and treatment

effects, such as biological markers, biomarkers, surrogate markers, surrogate endpoints, intermediate endpoints, and other terms that have overlapping meanings. This ambiguity stems from the involvement of a variety of disciplines (eg, clinical trialists, statisticians, regulators, and therapeutic developers) and different clinical research applications. To improve communication about this topic, an expert working group was convened by the National Institutes of Health and charged to propose terms, definitions, and a conceptual model.

DEFINITIONS

The following terms, definitions, and characteristics were proposed to describe biological measurements in therapeutic development and assessment.

Biological marker (biomarker): A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Biomarkers may have the greatest value in early efficacy and safety evaluations such as in vitro studies in tissue samples, in vivo studies in animal models, and early-phase clinical trials to establish "proof of concept." Biomarkers have many other valuable applications in disease detection and monitoring of health status. These applications include the following:

- use as a diagnostic tool for the identification of those patients with a disease or abnormal condition (eg, elevated blood glucose concentration for the diagnosis of diabetes mellitus)
- use as a tool for staging of disease (eg, measurements of carcinoembryonic antigen-125 for various cancers) or classification of the extent of disease (eg, prostate-specific antigen concentration in blood used to reflect extent of tumor growth and metastasis)
- use as an indicator of disease prognosis (eg, anatomic measurement of tumor shrinkage of certain cancers)
- use for prediction and monitoring of clinical response to an intervention (eg, blood cholesterol concentrations for determination of the risk of heart disease).

Clinical endpoint: A characteristic or variable that reflects how a patient feels, functions, or survives.

Clinical endpoints are distinct measurements or analyses of disease characteristics observed in a study or a clinical trial that reflect the effect of a therapeutic intervention. Clinical endpoints are the most credible characteristics used in the assessment of the benefits and risks of a therapeutic intervention in randomized clinical trials.

Surrogate endpoint: A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

Surrogate endpoints are a subset of biomarkers. Although all surrogate endpoints can be considered biomarkers, it is likely that only a few biomarkers will achieve surrogate endpoint status. The term *surrogate endpoint* applies primarily to endpoints in therapeutic intervention trials; however, it may sometimes apply in natural history or epidemiologic studies. It is important to point out that the same biomarkers used as surrogate endpoints in clinical trials are often extended to clinical practice in which disease responses are similarly measured. The use of biomarkers as surrogate endpoints in a clinical trial requires the specification of the clinical endpoints that are being substituted, class of therapeutic intervention being applied, and characteristics of population and disease state in which the substitution is being made. The term *surrogate* literally means "to substitute for"; therefore use of the term *surrogate marker* is discouraged because the term suggests that the substitution is for a marker rather than for a clinical endpoint.

The use of surrogate endpoints to establish therapeutic efficacy in registration trials is an established concept that has been addressed in regulation that enables the US Food and Drug Administration (FDA) to grant accelerated marketing approval for certain therapeutics.[20] This regulation states the following[20]:

> *Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity*. FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity. Approval under this section will be subject to the requirement that the applicant study the drug further, to verify and describe its clinical benefit, where there is uncertainty as to the relation of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome. Postmarketing studies would usually be studies already underway. When required to be conducted, such studies must also be adequate and well controlled. The applicant shall carry out any such studies with due diligence.[20]

A recent summary in which the status of surrogate endpoints in the clinical trials of cardiovascular therapies was

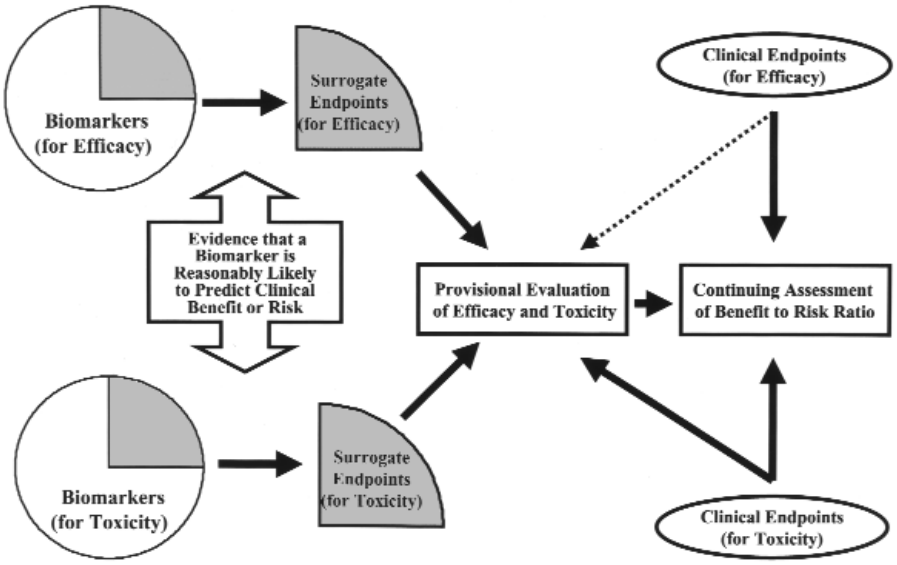

Fig 1. Conceptual model of the relationship of biomarkers, surrogate endpoints, and the process of evaluating therapeutic interventions.

addressed provided a perspective on the controversies regarding this strategy for regulatory approval.[21] The definitions provided here are consistent with this perspective and the use of the terms in the regulation.

CONCEPTUAL MODEL FOR BIOMARKERS AND SURROGATE ENDPOINTS

When the measurement of a biomarker is considered in the evaluation of a response to therapeutic intervention, it is important to identify the purpose that the biomarker serves in the drug development and evaluation process. A conceptual model was developed to show the relation of a biomarker to a clinical endpoint and the application of the biomarker as surrogate endpoint in the evaluation of therapeutic interventions (Fig 1). The model also shows that biomarkers may be useful in the assessment of safety, as well as efficacy.[22,23] Some biomarkers (eg, blood pressure) may have dual functions by assessing efficacy and safety.

A subset of biomarkers, represented by a quadrant of the biomarker circle, may achieve surrogate endpoint status. Characterization of a biomarker as a surrogate endpoint requires it to be "reasonably likely, based on epidemiologic, therapeutic, pathophysiologic or other evidence, to predict clinical benefit."[20] The utility of a biomarker as a surrogate endpoint requires demonstration of its accuracy (the correlation of the measure with the clinical endpoint) and precision (the reproducibility of the measure). Controlled clinical trials that evaluate the effect of a therapeutic intervention on a surrogate endpoint and provide evidence for its safety form the basis of a provisional evaluation of the intervention. When regulatory standards are met, approval of a new drug application may be based on this provisional evaluation. The term *provisional* indicates that this evaluation has not directly addressed the effect of the intervention on clinical outcomes and connotes an expectation that further evidence of the efficacy and safety of the therapeutic intervention will be acquired.

The assessment of clinical endpoints in controlled clinical trials provides the most convincing evidence for the benefit of an intervention. Information on safety may be derived both from controlled trials and from observational investigations in larger populations. The acquisition of evidence that directly supports the efficacy of a therapeutic intervention, together with the accrued information on safety, will then provide an assessment of benefit and risk, which is described in the FDA regulation[20] as ultimate outcome. For many, if not most, therapeutic interventions, the assessment of benefit and risk is not a one-time evaluation but rather a process that evolves with the accrual of knowledge about the actual consequences that derive from the therapeutic intervention.

EVALUATION OF THE LINKAGE OF THE BIOMARKER TO A CLINICAL ENDPOINT

In certain instances in which an intervention has been shown to have a benefit on a clinical endpoint, it may be of interest to evaluate retrospectively how well a surro-

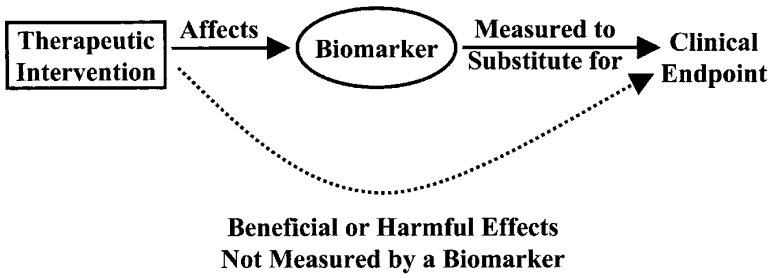

Fig 2. Effects of therapeutic interventions on biomarkers and clinical endpoints in clinical trials. In many circumstances, a therapeutic intervention will affect a clinical endpoint in a way that is not entirely accounted for by its effect on a biomarker. This is likely to occur in complex diseases in which a single biomarker may capture only a portion, or none, of the treatment effect. Interventions may also have unanticipated adverse consequences that diminish or completely offset the intended therapeutic benefits. The independent impact of these unanticipated beneficial or harmful effects of an intervention on clinical endpoints is represented by the *broken arrow.* Those biomarkers that do not account for a sufficient proportion of the treatment effect do not advance to surrogate endpoint status.

gate endpoint predicted that clinical outcome. The process of retrospectively linking a surrogate endpoint to a clinical endpoint has often been referred to as *validation*. This term is often used to reflect a significant statistical correlation; however, in this case it is considered that causal or mechanistic associations of the intervention with the disease process supports the consideration of a biomarker as a surrogate endpoint. In addition, the term *validation* is also often used to address performance characteristics (ie, sensitivity, specificity, and reproducibility) of a measurement or an assay technique. Importantly, a declaration that a surrogate endpoint is valid connotes that its validity is generalizable to include other interventions that affect the surrogate endpoint. Although such generalizations may be useful in the case of surrogate endpoints for certain diseases, such as HIV messenger ribonucleic acid viral load responses to highly-active antiretroviral therapy, there are other cases, such as the use of bone mineral density as a surrogate endpoint for osteoporosis therapies, for which there are limitations on the extent to which a correlation between surrogate endpoint and clinical endpoint for one intervention may be extrapolated to other classes of therapy. The use of the term *valid* implies that such limitations do not exist. For all of these reasons, the term *validation* is unsuitable for the description of the process of linking biomarkers to clinical endpoints, and the process of determining surrogate endpoint status is referred to as *evaluation*.

One approach used for the establishment of the linkage of a biomarker to a clinical endpoint is the estimation of proportion of treatment effect that is accounted for by the surrogate endpoint.[24,25] There are several ways to make this determination,[26] the most stringent of which requires a valid surrogate endpoint to account for all of the effects of the intervention on the clinical endpoint.[27] One approach used was the meta-analysis of treatment effects on the same biomarker in a series of trials in which different classes of therapies that affected the same clinical endpoint were used.[28,29]

The major concern about the use of biomarkers as surrogates for clinical endpoints is that in most circumstances not all treatment effects are fully accounted for by a single biomarker (Fig 2). The possible outcomes and concerns of reliance on a surrogate endpoint in decision making about efficacy and safety of therapeutic interventions have recently been reviewed.[11] Because evaluation of the qualities and characteristics of putative surrogate endpoints requires consideration of many factors, including the nature of the disease and the mechanism of action of the therapeutic intervention, it is not possible to describe uniform criteria that will apply to each application of a surrogate endpoint for each disease. Finally, the use of multiple biomarkers that represent various components of complex disease pathways may yield surrogate endpoints that offer a more comprehensive assessment of treatment effects.

CONCLUSIONS

Biomarkers serve a wide range of purposes in drug development, clinical trials, and therapeutic assessment strategies. Biomarkers can provide a basis for the selection of lead candidates for clinical trials, for contribution to the understanding of the pharmacology of candidates, and for characterization of the subtypes of disease for which a therapeutic intervention is most appropriate.

Given this scenario, there are minimal public health consequences of an inaccurate reliance on a biomarker.

Robust linkage of a biomarker with a clinical endpoint is not essential in early clinical development when the goal is confirmation of pharmacologic activity or optimization of dose regimens. Reliance on a biomarker for candidate selection entails the hazard that failure of a biomarker may lead to the elimination of potentially effective agents. On the other hand, substantial evidence that a biomarker will predict clinical benefit or risk is needed when use of the biomarker as a surrogate endpoint is proposed as the basis for regulatory approval. In this case, erroneous decisions based on invalid surrogate endpoints may have broad public health consequences.[11] As a result, attendant safeguards are provided that stipulate accelerated market withdrawal procedures for drug approval based on the accelerated approval provisions of the FDA.

The evaluation of disease intervention strategies can be facilitated and strengthened by the use of appropriate biomarkers that measure biological parameters of disease and therapeutic response in humans. The realization of the potential benefits that surrogate endpoints can bring in expediting of the development of safe and effective therapies will require an increased understanding of the linkage of biomarkers to clinical endpoints and will necessitate high levels of scientific scrutiny and rigor.

Developed by a working group of the National Institutes of Health Director's Initiative on Biomarkers and Surrogate Endpoints. Arthur J. Atkinson, Jr, MD, Warren G. Magnuson Clinical Center, National Institutes of Health, Bethesda, Md; Wayne A. Colburn, PhD, MDS Harris, Inc, Phoenix, Ariz; Victor G. DeGruttola, ScD, Harvard School of Public Health, Boston, Mass; David L. DeMets, PhD, Department of Biostatistics, University of Wisconsin, Madison, Wis; Gregory J. Downing, DO, PhD, Office of Science Policy National Institutes of Health, Bethesda, Md; Daniel F. Hoth, MD, Axys Pharmaceuticals, S San Francisco, Calif; John A. Oates, MD, Department of Medicine, Vanderbilt University, Nashville, Tenn; Carl C. Peck, MD, Center for Drug Development Science, Georgetown University, Washington, DC; Robert T. Schooley, MD, Department of Medicine, University of Colorado Health Sciences Center, Denver, Colo; Bert A. Spilker, PhD, MD, Pharmaceutical Research and Manufacturers of America, Washington, DC; Janet Woodcock, MD, Center for Drug Evaluation and Research, US Food and Drug Administration, Rockville, Md; and Scott L. Zeger, PhD, Department of Biostatistics, Johns Hopkins University School of Public Health and Hygiene, Baltimore, Md.

References

1. Carr G. A survey of the pharmaceutical industry. Economist 1998;346:1-18.
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.
3. Rolan P. The contribution of clinical pharmacology surrogates and models to drug development: a critical appraisal. Br J Clin Pharmacol 1997;44:219-25.
4. Blue JW, Colburn WA. Efficacy measures: surrogates or clinical outcomes. J Clin Pharmacol 1996;36:767-70.
5. Fowler JS, Volkow ND, Logan J, Wang G-J, MacGregor RR, Schlyer D, et al. Slow recovery of human brain MAO B after L-deprenyl (Selegeline) withdrawal. Synapse 1994;18:86-93.
6. Ciarmiello A, Del Vecchio S, Silvestro P, Potena AMI, Carriero MV, Thomas R, et al. Tumor clearance of technetium 99m-sestamibi as a predictor of response to neoadjuvant chemotherapy for locally advanced breast cancer. J Clin Oncol 1998;16:1677-83.
7. Lagakos SW, Hoth DF. Surrogate markers in AIDS: where are we? Where are we going? Ann Intern Med 1992;116:599-601.
8. Deyton L. Importance of surrogate markers in evaluation of antiviral therapy for HIV infection. JAMA 1996;276: 159-60.
9. Pozniak A. Surrogacy in HIV-1 clinical trials. Lancet 1997;351:536-7.
10. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, editors. Clinical measurement in drug evaluation. New York: John Wiley & Sons; 1995. p. 1-22.
11. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996;125: 605-13.
12. Echt DS, Liebson PR, Mitchell B, Peters RW, Obias-Manno D, Barker AH et al. Mortality and morbidity of patients receiving encainide, flecainide or placebo: the Cardiac Arrhythmia Suppression Trial. N Engl J Med 1991;324:781-8.
13. Psaty B, Weiss NS, Furberg CD, Koepsell TD, Siscovick DS, Rosendaal FR, et al. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. JAMA 1999; 282:786-90.
14. SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: Final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA 1991;265:3255-64.
15. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival study (4S). Lancet 1994;344:1383-9.
16. Ridker PM. Evaluating novel cardiovascular risk factors: can we better predict heart attacks? Ann Intern Med 1999;130:933-7.
17. Ridker PM, Rifai N, Pfeffer MA, Sacks F, Braunwald E, for the Cholesterol and Recurrent Events (CARE) Investigators. Long-term effects of pravastatin on plasma concentration of C-reactive protein. Circulation 1999;100:230-5.
18. The International Chronic Granulomatous Diseases Cooperative Study Group. A controlled trial of interferon

gamma to prevent infection in chronic granulomatous disease. N Engl J Med 1991;324:509-16.
19. Boissel JP, Collet JP, Moleur P, Haugh M. Surrogate endpoints: a basis for a rational approach. Eur J Clin Pharmacol 1992;343:235-44.
20. The Food and Drug Modernization Act of 1997. Title 21 Code of Federal Regulations Part 314 Subpart H Section 314.500.
21. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? JAMA 1999;282:790-5.
22. MacGregor JT, Farr S, Tucker JD, Heddle JA, Tice RR, Turteltaub KW. New molecular endpoints and methods for routine toxicity testing. Fundam Appl Toxicol 1995; 26:156-73.
23. Santella RM. DNA damage as an intermediate biomarker in intervention studies. Proc Soc Exp Biol Med 1997; 216:166-71.
24. Lin DY, Fleming TR, DeGruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med 1997;16:1515-27.
25. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 1992;11:167-78.
26. Fleming TR, DeGruttola V, DeMets DL. Surrogate Endpoints. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. Vol 6. New York: John Wiley & Sons; 1998. p. 4425-31.
27. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431-40.
28. Hughes MD, DeGruttola V, Wells SL. Evaluating surrogate markers. J Acquir Immune Defic Syndr 1995;10:S1-8.
29. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med 1997;16: 1965-82.