Journal of Cellular Biochemistry, Supplement 16G:27-32 (1992)

Statistical Validation of Intermediate Markers of Precancer for Use as Endpoints in Chemoprevention Trials

Laurence S. Freedman[1], Arthur Schatzkin[2], and Mark H. Schiffman[3]

[1]Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Executive Plaza North, Suite 344, Bethesda, Maryland 20892
[2]Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Maryland 20892
[3]Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20892

Abstract Using an intermediate marker of precancer as an endpoint for evaluating agents that may prevent cancer involves a presumption that the modification of the marker will be accompanied by a modification of cancer incidence. This presumption can hold only if the marker is on or very closely linked to a causal pathway. Epidemiologists have discussed the nature of evidence required to infer causal relationships, and we briefly survey their work. Studies relating exposure (E) to marker (M) provide only indirect evidence for causality. Those relating marker (M) to disease (D) are more relevant. We propose a new validation criterion based on an analysis of the three-way relationship of exposure (E), marker (M) and disease (D). We discuss the level of evidence required for using intermediate markers as endpoints for Phase II and Phase III trials, and propose very stringent criteria for Phase III trials. For Phase II trials, we propose less stringent criteria, but still recommend that the marker (M) should have been shown to have a strong association with disease (D).

Key words: causal relationships, chemoprevention, intermediate biomarker, marker validation, Phase II trials, Phase III trials

Intermediate biomarkers of precancer have the potential for use in measuring exposure to possibly carcinogenic agents, in identifying populations at high risk of developing cancer, or as surrogate endpoints for cancer in prevention trials. In this paper we discuss the last of these three uses. The attraction of using a biomarker as an endpoint for a cancer prevention trial is in the saving of time and money. Since any effect of a cancer prevention agent on the biomarker will precede the effect on cancer, the duration of trials will thereby be shortened, perhaps by several years. Furthermore, since biomarker change is more likely to occur in a larger proportion of subjects than will cancer diagnosis, the trial will require fewer subjects, perhaps to the extent of reducing sample size from thousands to hundreds. Thus the benefits to cancer prevention research of finding a reliable cancer surrogate could be enormous.

However, it is difficult to establish that a biomarker is a valid cancer surrogate. We need to establish that the marker lies on (or is very closely linked to) a causal pathway to the cancer, because only then will we feel confident that when modifying the marker we are also modifying cancer incidence. Thus the notion of causal relationship is central to the use of a surrogate endpoint. In the next section we review very briefly the contributions from epidemiologists to the question of inferring causality. It should be mentioned that the discovery of a new marker that is firmly established to be on the causal pathway of a cancer is a rare event in cancer research. The implications of such a discovery would go far beyond economics in cancer prevention research, and would probably lead to new theories of etiology and new modes of prevention.

CAUSALITY IN EPIDEMIOLOGY

For a good review of the epidemiologist's pragmatic approach to causality, the reader is referred to the recent paper "What Is A Cause and How Do We Know One? A

Published 1992 Wiley-Liss, Inc.

Grammar for Pragmatic Epidemiology" by Mervyn Susser [1]. Susser describes the 1964 Report of the Advisory Committee to the US Surgeon General "Smoking and Health" [2] as a public health landmark for the present era of chronic disease epidemiology. In this report, five criteria were given for inferring causality from a given association between a risk factor and disease. These were strength, specificity, consistency, time order, and coherence. In terms of a marker (M) and disease (D) these five criteria would require that:

(a) the association between M and D is observed to be strong;

(b) the association is specific, i.e. M is not related simultaneously to many other diseases D_1, D_2, D_3, ..., or M is not a series of many markers M_1, M_2, M_3, ... associated with D;

(c) the association is consistently observed over several studies;

(d) there is an established time order between changes in M and changes in D; for example, the studies showing the association are prospective, meaning that M is assessed at baseline and subjects are followed to determine if D develops;

(e) the association is coherent with other biological evidence suggesting a causal relationship between M and D.

These five criteria were elaborated and criticized by Hill [3] and Susser [4] among others. In the context of markers the specificity requirement does not appear particularly persuasive and the coherence requirement is a sine qua non (but see the next section). However, strength, consistency and time order are all relevant. In general the work reviewed above emphasizes the importance of examining the direct association between marker and disease in evaluating whether the marker is causally related to disease.

EXPOSURE, MARKER AND DISEASE

The epidemiologists' work on causation focused mainly on relationships between risk factor or exposure (e.g., smoking) and disease. However biomarkers have generally an intermediate role, with modification or occurrence of biomarker events happening between the time of one or more exposure(s) and the time of clinical diagnosis of the disease. We believe that in some circumstances a particularly strong form of evidence for causality can be obtained by analyzing the three-way relationship between exposure (E), marker (M) and disease (D). We will call studies in which all these components are assessed E–M–D studies.

A review of the literature on biomarkers will reveal that such studies are currently rarely performed. As an example, consider the review by Boyd and McGuire [5] on the relationship between HDL-cholesterol levels and breast cancer incidence. The studies that have been reported may be divided into two classes (Table I). First we have the studies that relate the marker (HDL-cholesterol) to various exposures (putative risk factors for breast cancer). In line with the notation we have introduced, we call these E–M studies. In Boyd and McGuire's review are included several of these studies. For example in three studies, HDL-cholesterol levels had been compared in groups of women from countries with widely varying breast cancer incidence rates. Other investigators had related HDL-cholesterol levels to body weight, to alcohol intake and to pregnancy. In ten intervention studies HDL-cholesterol levels were assessed before and after the adoption of a diet low in fat.

A second set of studies were concerned with relating HDL-cholesterol to breast cancer incidence. We call these M–D studies. In Boyd and McGuire's review only retrospective case-control studies were described: HDL-cholesterol measurements were performed in women with a breast cancer diagnosis and control women with no breast cancer diagnosis. Eight such studies were reported and the results were conflicting. No prospective cohort studies, in which HDL-cholesterol was measured at baseline and related to later breast diagnosis, were reported. In addition, no studies of exposure, marker and disease (E–M–D) were reported.

TABLE I. Classification of Studies Pertaining to the Relationship Between HDL-Cholesterol and Breast Cancer, Cited by Boyd and McGuire [5].
[Number of studies cited in parentheses.]

E–M Studies (Exposures (E) listed below: Marker (M) is **HDL-Cholesterol)**

- Country of residence (3)
- Pregnancy (1)
- Dietary fat intake (experimental studies) (10)
- Dietary fat intake (observational studies) (4)
- Alcohol intake (4)
- Body weight (1)
- Endogenous hormone levels (9)
- Hormone replacement therapy (4)
- Socioeconomic status (1)

M–D Studies (Marker (M) is HDL-Cholesterol: Disease **(D) is Breast Cancer)**

- Case-control studies (8)
- No prospective cohort studies

E–M–D Studies

- None

The situation revealed by this review is likely to be typical of the current status of research into many potential biomarkers. In regard to the question of causality, the E–M studies provide only indirect evidence since they cannot by themselves establish a direct association between marker and disease. To show that low HDL-cholesterol is associated with previous pregnancy may be suggestive of a possible link with breast cancer, but cannot substitute for the demonstration of a direct association. Similarly, showing that HDL-cholesterol may be modified by changing the amount of fat in the diet constitutes strong evidence neither for a direct link between HDL-cholesterol and breast cancer, nor for the fat-breast cancer hypothesis.

The M–D case-control studies in this review correspond to classic risk factor-disease studies that are the bread and butter of epidemiologists' work. In the biomarker context, retrospective case-control studies will often, but not always, be open to question because of the possibility of reverse causation, i.e. the disease causing the change in the marker. In other words case-control studies do not meet the requirement to show time order in the association. This is less important when evidence for the time order is available elsewhere, but in the case of HDL-cholesterol and breast cancer such evidence is absent. Prospective studies will often be required, therefore, for establishing marker-disease relationships.

Although E–M–D studies are rarely performed currently, they afford a particularly strong type of validation for biomarkers. As explained in Schatzkin et al. [6], E–M–D studies allow us to address the question of whether a marker mediates the effect of exposure (or intervention) on disease. Such mediation would imply that, knowing the effect of the exposure upon the marker, we would be able to predict its effect upon disease. Of course, if we establish mediation of a marker with respect to one exposure there is no guarantee that the same marker will mediate another exposure, as there may exist several different causal pathways to the disease. However, the marker would nevertheless become an important intermediate endpoint since establishing mediation for one known exposure would imply that an intervention that modifies the marker would, in the absence of simultaneously affecting a second pathway, change the disease incidence. The strongest type of validation would show that all the known risk factors for a certain disease are mediated by the same marker.

Table II. Odds-Ratio of Cervical Dysplasia according to Number of Sexual Partners unadjusted and adjusted for HPV

	Number of sexual partners				
	1	2	3-5	6-9	10+
Unadjusted	1.0	1.7	3.1*	4.7*	4.4*
Adjusted for HPV status	1.0	1.0	1.1	1.5	1.6

* $P < 0.001$: Estimated Effects are more than 4 times their standard errors
Odds ratio without an asterisk are not significant at 5% level.

One may view E-M-D validation as a demonstration of the coherence criterion for causality, where we are extending the definition of coherence to include coherence with other epidemiologic evidence. That is, having already established that E is a risk factor for D, we are investigating whether M's relationships with E and D are coherent with its hypothesized role as a mediating factor.

The statistical analysis for validation of a marker in an E–M–D study requires estimation of the exposure effect upon disease, first unadjusted, and second adjusted for marker status. Under mediation, one would expect the unadjusted analysis to show a strong highly significant relationship between exposure and disease (with the estimated effect perhaps 4 or more times its standard error) and the adjusted analysis to show this relationship to disappear (or at least become statistically non-significant).

Although the discovery of such strong evidence is unusual, recently available data provide a striking example. In an NIH-sponsored case-control study of human papillomavirus (HPV) infection and cervical dysplasia conducted in Portland, Oregon, women with cervical dysplasia and controls were questioned regarding conventional risk factors (such as sexual activity) and were also tested for HPV infection using a new, highly accurate polymerase chain reaction technique. Data from this study are shown in Table II. It can be seen that, while the unadjusted odds ratios for women with 3 or more sexual partners are large and highly significant, the corresponding odds ratios adjusted for HPV status are close to 1 and are not statistically significant. These data make a strong case for regarding HPV status as a valid intermediate biomarker mediating between an established risk factor (sexual activity) for cervical dysplasia and the development of dysplasia. Further analysis of the data, particularly separating high-grade and low-grade dysplasia, is important and will be presented elsewhere.

In this section we have laid out the full range of types of study and strengths of evidence available for validating markers. The E-M-D validation is the 'cadillac' version of biomarker validation and cannot be expected as a regular occurrence. Other weaker types of validation can still be useful. In the next section we discuss how the considerations of this section may be applied to the use of biomarkers in chemoprevention trials.

APPLICATION TO CHEMOPREVENTION TRIALS

Following pre-clinical testing, cancer chemoprevention agents are tested in humans in three phases. In Phase I trials

the safety of an agent is tested, and, if it passes this test, the efficacy of the agent is subjected to preliminary evaluation in Phase II trials. If the agent appears sufficiently promising in these early phase studies then one or more Phase III trials will be conducted to provide a definitive evaluation of the agent's use in cancer chemoprevention.

Typically Phase II trials are conducted using biomarkers as endpoints, whereas Phase III trials employ cancer incidence as the endpoint. For example, in a Phase II trial of piroxicam as an agent for preventing colon cancer, fewer than 100 subjects were assessed with regard to changes in prostaglandin E2 and prostaglandin synthetase levels. In a Phase III trial of beta-carotene, approximately 22,000 US physicians are being followed, with death from cancer as a main endpoint.

Two questions arise. Firstly are there criteria we can use for deciding whether a biomarker is sufficiently validated for use in a Phase II trial? Secondly, could we ever use biomarkers in place of a cancer endpoint in a Phase III trial? (As mentioned in the Introduction, the savings could be enormous.)

The second question is the easier one to answer. Since in Phase III we require a definitive evaluation, we can only use a biomarker that we feel sure is on the causal pathway. Thus the evidence must be of the most stringent nature, and in our opinion the marker should, where possible, be validated in an E–M–D study using the statistical analysis described in the section on Exposure, Marker and Disease, where the exposure E is thought to operate along the same pathway as the chemoprevention agent being tested. Failing this, the consistent demonstration in several M–D studies of a strong relationship between marker and disease, and particularly a high attributable proportion [6], when coupled with laboratory evidence on the causal nature of the relationship, might be regarded as acceptable.

The questions regarding criteria for the use of biomarkers in Phase II trial are more difficult because the aim of a Phase II trial is less well-defined. Phase II is intended to act as a final screen for the most promising of the chemoprevention agents to determine which agents should be given the definitive test. Thus the criteria for biomarkers in Phase II need not be as strict as for their use in Phase III trials. On the other hand it is desirable to use biomarkers that are going to be good predictors of the true efficacy of the agents. We suggest that there should be at least some evidence of a direct association between the biomarker and cancer incidence. As a compromise, one would not demand that the evidence be prospective in nature, and evidence from case-control studies would be allowed. In the event of several such studies having been conducted, some consistency of results should be required. In summary we suggest that for use as a Phase II study endpoint, a biomarker should be demonstrated to have an association with the cancer incidence and that the estimated attributable proportion be at least moderate (usually 0.25 or larger), although the temporality of the relationship might not yet have been shown. Associations of the biomarker with risk factors for the cancer should not by themselves persuade investigators to use the biomarker as an endpoint. In particular, it is important to avoid the circular logic that if a putatively chemopreventive agent modifies a biomarker, then the marker is a suitable endpoint for a Phase II trial.

AN EXAMPLE: CALCIUM AND CELL PROLIFERATION RATE

The study of calcium as a potential agent for colorectal cancer prevention provides an example of the use of a biomarker in a Phase II chemoprevention trial. Lipkin and Newmark [7] have described a recent trial in which they demonstrate that calcium supplementation reduces the proliferative activity of colonic epithelial cells in subjects at high risk of familial colonic cancer. The study is a good example of a Phase II trial, using cell proliferation rate as an endpoint.

A later report by Lipkin et al [8] provides evidence of an association between cell proliferation rate and cancer (M-D study). In this study cell proliferation rates were shown to be increased among patients with colorectal cancer compared to normal subjects, using a case-control design. Moreover the reported differences between

cases and controls were large enough to be consistent with a high attributable proportion for the cell proliferation rate. It would be important to rule out the possibility that the presence of malignancy influences the marker, but at present no prospective studies relating cell proliferation rate and colorectal cancer incidence have been conducted. However, there are two prospective studies under way that are investigating the relationship between cell proliferation rate and recurrence of adenomatous polyps, a precancerous lesion.

This level of evidence for cell proliferation rate as an intermediate marker of precancer happens to meet closely our suggested minimum requirements for a Phase II trial study endpoint. It would be interesting to analyze the evidence for several other cellular, biochemical and genetic biomarkers (eg., see Lippman et al [9]) using the criteria we have suggested. Such an exercise would help to bring the concerns of epidemiologists and laboratory scientists in closer harmony, the main goal being to develop more quickly effective cancer prevention agents.

ACKNOWLEDGMENTS

We thank Regina Mease for excellent secretarial assistance.

REFERENCES

1. Susser M: What is a cause and how do we know one? A grammar for pragmatic epidemiology. Am J Epidemiol 133:635-648, 1991.

2. Smoking and Health. Report of the Advisory Committee to the Surgeon General. Wash, DC: US Dept of Health, Education and Welfare, 1964.

3. Hill AB: Environmental and disease: association or causation? Proc Royal Soc Medicine 58:295-300, 1965.

4. Susser M: Causal thinking in the health sciences: concepts and strategies in epidemiology. New York: Oxford University Press, 1973.

5. Boyd NF, McGuire V: Evidence of association between plasma high-density lipoprotein cholesterol and risk factors for breast cancer. J Natl Cancer Inst 82:460-468, 1990.

6. Schatzkin A, Freedman LS, Schiffman MH, Dawsey SM: Validation of intermediate end points in cancer research. J Natl Cancer Inst 82:1746-1752, 1990.

7. Lipkin M, Newmark H: Effect of added dietary calcium on colonic epithelial cell proliferation in subjects at high risk for familial colonic cancer. New Engl. J. Med. 313: 1381-1384, 1985.

8. Lipkin M, Enker W E, Winawer S J: Tritiated-thymidine labeling of rectal epithelial cells in 'non-prep' biopsies of individuals at increased risk for colonic neoplasia. Cancer Letters 37: 153-161, 1987.

9. Lippman SM, Lee JS, Lotan R, Hittelman W, Wargovich MJ, Hong WK: Biomarkers as intermediate endpoints in chemoprevention trials. J Natl Cancer Inst 82: 555-560, 1990.