

Editorial**Surrogate End Points in Cancer Research: A Critique**

Arthur Schatzkin,¹ Laurence S. Freedman,
Joanne Dorgan, Lisa M. McShane, Mark H. Schiffman,
and Sanford M. Dawsey

Department of Health and Human Services, National Cancer Institute, NIH,
Bethesda, Maryland 20892

Abstract

Studies using surrogate end points of malignant disease may be smaller, shorter, and less expensive than studies with incident cancer end points. Researchers have proposed a broad range of histological, cellular, and molecular markers as surrogate end points for cancer (SECs). We define a valid SEC as follows: the effect of an intervention on (or the association of a risk factor with) the SEC is concordant with its effect on (or association with) incident cancer. Adenomatous polyps and persistent human papillomavirus infections are examples of reasonably valid SECs (for colorectal and cervical cancer, respectively) because these markers are necessary precursors of most of these malignancies. Inferences from other potential SECs, however, are problematic if there exist major alternative causal pathways to malignancy bypassing the SEC. Furthermore, in such circumstances, an SEC that is valid for one intervention or exposure may not be valid for another. Even for those end points without such major alternative pathways, an intervention could differentially affect two intermediate markers on the same pathway, thus disturbing the concordance between its effect on a given SEC and its effect on cancer. Thus, an understanding of the causal structure underlying the relations of interventions/exposures, potential SECs, and cancer is critical in evaluating SECs. Three questions are pertinent to elucidating this structure: (a) What is the relation of the SEC to cancer? (b) What is the relation of the intervention/exposure to the SEC? and (c) To what extent does the SEC mediate the relation between the intervention/exposure and cancer? Ecological, metabolic, observational epidemiological, and intervention studies may provide data relevant to one or more of these questions. Data on SEC variability are critical in evaluating whether marker findings have been attenuated by random sources of intra-individual variation. We emphasize the importance of conducting studies, especially SEC-cancer and intervention/exposure-SEC-cancer mediation studies, to evaluate problematic SECs such as epithelial cell hyperproliferation. For some time to come, hard and

policy-relevant evidence on cancer etiology and prevention will emerge only from studies with cancer end points or, at a somewhat lower level of certainty, SECs that are (for the most part) obligatory steps on the causal pathway to malignant disease.

Even our most common cancers occur relatively infrequently. The age-adjusted annual incidence of breast cancer among United States women is approximately 100 per 100,000, or 0.1%. For colorectal cancer, the incidence among men and women combined is a little under 50 per 100,000, or 0.05%. Because the diagnosis of cancer is such a relatively rare event in the population, clinical trials or observational epidemiological studies with incident cancer end points must be very large, very lengthy, or both, which generally means very expensive as well. Studies using surrogate end points of malignant disease can be smaller, faster, and cheaper than studies with incident cancer end points. It is not surprising, then, that cancer researchers have long been interested in using these markers.

It is by no means certain, however, that studies using such surrogates give us the right answers about cancer *per se*. This paper presents a theoretical framework and research strategy for evaluating whether results of surrogate studies are generalizable to malignant disease.

Definition of a Surrogate End Point Marker for Cancer

We define a SEC² as follows: a surrogate for incident cancer yields a valid test of the null hypothesis of no association between intervention and incident cancer. In other words, the effect of an intervention on the SEC is concordant with its effect on cancer incidence, or, for observational epidemiological studies, the association of an exposure with the SEC is concordant with its association with cancer incidence. ("Concordance" implies proportional effects. If, for example, a large change in the SEC means a large change in cancer incidence, then a small change in the SEC would mean a small change in cancer incidence.) If the SEC meets these conditions, we will call it a "valid" surrogate for that cancer. These conditions follow from the criteria proposed by Prentice (1).

Examples of SECs

SECs may comprise a broad range of biological phenomena. Histological changes, involving both cellular and architectural abnormalities (structural or functional) and including both neoplastic and nonneoplastic lesions, are potential SECs. Adenomatous polyps of the large bowel (2) or cervical intraepithelial neoplasia (3) constitute neoplastic examples. Nonneoplastic lesions suggested as SECs include bronchial metaplasia for

Received 5/1/96; revised 9/6/96; accepted 9/12/96.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ To whom requests for reprints should be addressed, at Department of Health and Human Services, National Cancer Institute, NIH, 9000 Rockville Pike, EPN 211, Bethesda, MD 20892. Phone: (301) 496-8559; Fax: (301) 402-0553.

² The abbreviations used are: SEC, surrogate end point for cancer; HPV, human papillomavirus; AP, attributable proportion; RR, relative risk; ICC, intra-class correlation coefficient.



Pathway to cancer (CA) with single exposure (E1) and single marker

cancer (4), atrophic gastritis and intestinal metaplasia for cancer (5), and leukoplakia for oral cancer (6). Cellular phenomena, including proliferation (7) or apoptosis (8), may prove to be SECs. Potential molecular SECs, such as *ras* and other sporadic gene mutations (9), hypomethylation (10), chemical-DNA adducts (11), or mutagen sensitivity (12) involve structural or functional alterations of chromosomal DNA.

SECs may also include infectious processes, such as HPV infection of cervical epithelial tissue (13) and *Helicobacter pylori* infection of gastric tissue (14). Another set of possible SECs are tissue changes defined on the basis of various imaging techniques, such as mammographic parenchymal patterns (15) or ovarian ultrasound abnormalities (16). Finally, blood or tissue levels of bioactive substances, regarded as risk factors in the traditional epidemiological approach, may also be surrogate end points for certain malignancies. Examples include blood levels of various steroid hormones for malignancies of the breast, ovary, or prostate (17). Other SECs that entail structural or functional changes in target tissue from which cancer arises, levels of bioactive substances like hormones may be assayed in blood or tissue distal from the target tissue.

Quantifying SEC Validity: Logical Considerations

The validity of a potential SEC depends on the causal interrelationships of intervention/exposure, SEC, other markers, and cancer.

The simplest causal pathway involving a potential SEC is shown in Fig. 1. E1 represents an environmental or host factor. A change in E1 would alter SEC positivity and thereby modify the incidence of cancer. The SEC, by our definition, is a valid surrogate for cancer.

The causal pathway reflected in Fig. 1 rarely (if ever) occurs in nature. Figs. 2 and 3 depict more complex and realistic pictures of carcinogenesis.

In Fig. 2, E1 influences cancer through two alternative pathways, one through the potential SEC, the other through another marker (MARKER2). To the extent that E1 works through the alternative MARKER2 pathway (meaning that the SEC is not necessary for cancer), we cannot be certain that SEC is a valid surrogate in studies involving E1. This is because E1 may affect MARKER2 in a way that offsets its influence on the SEC; the final effect on cancer is unknown. If, for example, E1 increases SEC positivity but increases MARKER2 positivity, E1 would increase cancer incidence.

In Fig. 3, the joint action of two markers [the potential SEC and some other marker (MARKER2)] is necessary for the development of cancer. E1 may affect SEC or MARKER2. Again, we cannot be certain that SEC is a valid cancer surrogate in studies of E1 because E1 may affect SEC and MARKER2 in offsetting ways.

Figs. 2 and 3, although more complex than Fig. 1, represent idealized and simplified scenarios in their own right. Given the known cascades of growth promotion and inhibition that characterize cell biology, one can easily imagine still more complex situations involving combinations of pathways reflected in Figs. 2 and 3.

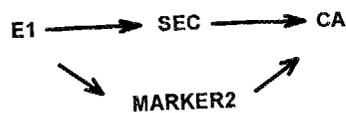


Fig. 2. Single exposure (E1) with pathways to cancer (CA) through two alternative markers (SEC and MARKER2).



Fig. 3. Pathway to cancer (CA) from exposure (E1) requires joint action of two markers (SEC and MARKER2).

Illustrations of These Logical Considerations

Consider large bowel adenomatous polyps, used increasingly as a surrogate for colorectal cancer. In Fig. 4 (pathway *a*), we postulate an event *X* that is necessary for an adenomatous polyp to progress to colorectal cancer. Two types of polypoid adenomas exist, those without *X* (innocent adenomas not progressing to cancer) and those with *X* (bad adenomas progressing to cancer). Both types are observable but indistinguishable through a colonoscope. Furthermore, there exist flat areas of dysplasia with *X* (not observable through the colonoscope) that also progress to cancer (18).

Suppose we have an intervention (a low-fat eating plan, for example), that reduces E1 (say, some mutagenic fecal constituent) and thereby diminishes the pool of adenomas susceptible to the relatively rare *X* events. Unless E1 increases the rate of *X* events in this reduced adenoma pool (an unlikely but theoretically possible scenario, reflected in Fig. 3), this intervention necessarily reduces the number of bad adenomas and therefore lowers the incidence of colorectal cancer.

The existence of the flat dysplasia pathway complicates things. Our intervention has no effect on pathway *b*, although it reduces observed adenomas via pathway *a*. To the extent that pathway *b* contributes to colorectal carcinogenesis, adenoma development may not be a valid SEC for cancer. As pathway *a* becomes the less common of the two routes to cancer, an investigator observing fewer adenomas developing among intervention participants could conclude that the intervention reduces colorectal cancer incidence, when in fact the intervention might have a quite different effect on cancer occurrence through pathway *b*. A large body of evidence, however, suggests that most colorectal cancers do develop through pathway *a* (the adenoma-carcinoma sequence; Refs. 18 and 19). Therefore, an intervention reducing adenomatous polyp recurrence would likely reduce colorectal cancer incidence. Adenoma recurrence is a reasonably valid SEC.

HPV infection in cervical cancer seems to be analogous to the adenoma-colorectal cancer example. The overwhelmingly large proportion of cervical cancer requires prior HPV infection (13). The relatively rare *X* event is whatever (still unknown) immunological deficit leads to persistent HPV infection. HPV persistence results in inactivation, by the E6 and E7 proteins of the HPV genome, of p53 and pRb tumor suppressor genes, leading in turn to increasingly severe intraepithelial neoplasia and, eventually, cancer. It is currently thought, however, that a small proportion of cervical cancer can arise as a result of tumor suppressor gene product inactivation occurring by mutation in the absence of HPV infection. Because most cervical cancer

ARCH

ast Cancer
ada
ngle Park, NC

cer

ancer

mit abstracts

1997

ancer Research
816
ill West
3
3 (FAX)

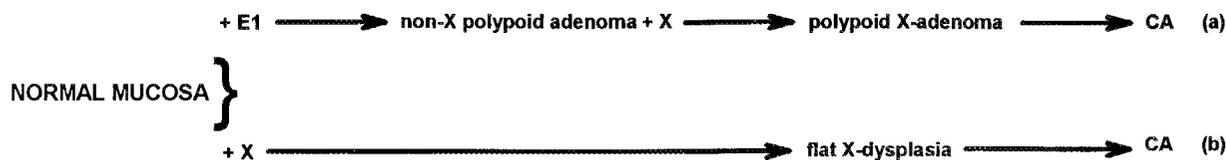


Fig. 4. Alternative pathways, with (a) and without (b) an adenomatous polyp step, to colorectal cancer. *E1*, exposure; *X*, event necessary for development of cancer.

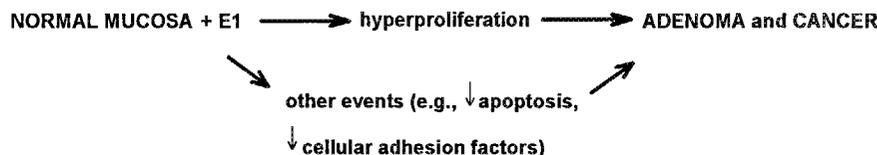


Fig. 5. Alternative pathways, with and without mucosal hyperproliferation, to colorectal cancer. *E1*, exposure.

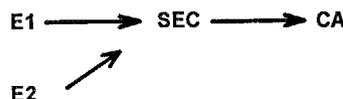


Fig. 6. Two exposures (*E1* and *E2*) lead to cancer (*CA*) through a single marker (*SEC*).

does occur through HPV infection, an intervention that eliminates or reduces HPV infection would probably decrease cervical cancer incidence.

A third example in this vein is Barrett's esophagus, a metaplastic change from squamous to columnar epithelium in the lower esophagus that is thought to be a necessary precursor to most cases of esophageal adenocarcinoma (20). Although gastric acid reflux is the primary precipitant of this metaplastic change, other factors may be required in the transition from Barrett's epithelium through dysplasia to adenocarcinoma. Although a small proportion of esophageal adenocarcinomas are known to arise from esophageal submucosal glands, independent of the Barrett's epithelium pathway, an intervention [photoblation (20) or electrocoagulation (21)] that eliminates the Barrett's epithelium would likely greatly decrease the incidence of esophageal adenocarcinoma.

Inferences to cancer from other potential SECs, however, are considerably more problematic. Fig. 5 depicts plausible causal pathways involving colorectal epithelial cell proliferation. (The adenoma step depicted in Fig. 4 is simplified here.) *E1* is again an exposure amenable to modification. To the extent that other cellular or molecular events, such as diminished apoptosis or altered cellular adhesion factors, constitute an important causal pathway from *E1* to cancer, we cannot be sure that the relation of *E1* to these other events does not offset the effect of *E1* on cancer through proliferation. Cell proliferation is a problematic SEC because the relative importance of the alternative pathways (through events other than proliferation) is simply unknown.

Exposure Dependence

In Fig. 6 we return to the simple idealized scheme from Fig. 1, but now we add another exposure, *E2*. Both *E1* and *E2* in Fig. 6 work through a single SEC on the path to cancer. Because SEC is a necessary precursor for cancer, the validity of this SEC

is exposure-independent. That is, any other exposure (*E2*) that influences cancer must operate through the SEC; the SEC is valid for studies of *E2* as well as those of *E1*.

In Fig. 7, we consider what happens when *E2* enters into the more complex scenario depicted in Fig. 2. In Fig. 7, the existence of a nontrivial alternative pathway (through MARKER2) means that the validity of the SEC is exposure-dependent. Even if *E1* works primarily through SEC and affects MARKER2 minimally, suggesting that SEC is reasonably valid for *E1*-cancer studies, we cannot assume that the *E2*-MARKER2-cancer pathway plays a similarly minor role in the development of cancer. A parallel argument holds for the multiple-marker scenario depicted in Fig. 3: a SEC valid for studies of *E1* and cancer may not be valid for another intervention (*E2*) that affects MARKER2 in ways that counterbalance its effect on the SEC.

A given agent, for example, might influence colorectal carcinogenesis largely through its influence on cell proliferation (Fig. 5). In that case, cell proliferation is a reasonably valid SEC for the first agent vis-à-vis colorectal cancer. A second agent, however, might affect cell proliferation very little (or not at all) but could increase apoptosis sufficiently to decrease cancer incidence. Focusing only on cell proliferation would give a falsely pessimistic impression of the efficacy of the second agent in inhibiting colorectal carcinogenesis.

Investigating Causal Pathways Involving SECs

We have argued that the causal structure underlying the relations of interventions/exposures, potential SECs, and cancer is critical in evaluating such SECs (22). Data helpful in revealing this structure can emerge from investigations of three questions: (a) Is the potential SEC associated with cancer (in particular, how large is the attributable proportion)? (b) Is the intervention/exposure associated with the potential SEC? and (c) Does the potential SEC mediate the relation of the intervention/exposure to cancer?

Traditional epidemiological parameters are useful in carrying out these investigations. For simplicity, we refer in the following discussion to potential SECs that are either positive or negative. The arguments offered here, however, may be extended to encompass markers measured as continuous variables.

Results of an epidemiological study of a potential SEC in relation to cancer are depicted generally in Fig. 8. If this were



Fig. 7. Each of two exposures ($E1$ and $E2$) leads to cancer (CA) through two markers (SEC and $MARKER2$).

an observational cohort study, the association between SEC and cancer would be indicated by the RR, defined as $[a/(a + b)]/[c/(c + d)]$. For a case-control study, the RR would be estimated by the odds ratio, defined as ad/bc . A RR (or odds ratio) of 1.0 indicates no association between the potential SEC and cancer. AP is an epidemiological measure that indicates the proportion of cancer that is attributable to SEC positivity. $AP = S(1 - 1/R)$, where $R = RR$ and $S =$ sensitivity, defined as $a/(a + c)$. An AP of 1.0 means that marker positivity is necessary for the development of cancer; that is, the carcinogenic pathway must go through this positive marker.

Question 1: SEC-Cancer

Clearly, for a biomarker to be a reasonable SEC, it must have something to do with the incidence of cancer. We would like to quantify the extent to which marker status correlates with or predicts the incidence of cancer.

Observational epidemiological studies are important vehicles for examining this SEC-cancer question. In a recent case-control study, Schiffman *et al.* showed a markedly increased risk of severe cervical neoplasia for those with HPV infection (23). Toniolo *et al.*, in a case-control study nested within a prospective cohort, observed a direct relation between serum estrogens and subsequent breast cancer (24). (Case-control studies nested in prospective designs have the virtue of avoiding reverse causation, whereby cancer affects marker status.) Observational cohort studies may also be nested in trials. In the Polyp Prevention Trial (25), for example, it will be possible to relate baseline proliferation indices to subsequent adenoma recurrence. (We have referred here to studies with neoplastic cancer precursor end points, such as cervical intraepithelial neoplasia and adenomas. For purposes of discussion, here we consider these as proxies for cancer, although, as we have shown, the validity of these precursor end points is not absolute.)

Ecological studies may provide pertinent (if indirect) information on the SEC-cancer question. Researchers have examined, for example, mean proliferative indices in groups at different risks of colorectal cancer (26). In ecological studies, as opposed to observational studies with both marker and disease information on individuals, the link between marker and disease is indirect; one cannot be certain that those who are marker-positive are the ones with increased incidence of cancer.

The AP is of great value here in evaluating the importance of alternative pathways (22). In idealized Fig. 1, the AP for SEC is 1.0. In the more realistic Fig. 2, however, with at least two pathways to cancer, the AP is <1.0 . If the AP for SEC is high, however, even if <1.0 , it suggests that the alternative MARKER 2 pathway plays a small role in the development of cancer. An AP substantially lower than 1.0 suggests that one or more alternative pathways is indeed operative.

In addition to epidemiological investigations of the SEC-cancer question, pathological, cell, and molecular biological

		CANCER	
		+	-
SEC	+	a	b
	-	c	d

Fig. 8. General relation between SEC and cancer.

studies may yield pertinent information. The adenoma-carcinoma sequence has received support from studies showing carcinomatous foci in adenomas and adenomatous foci within carcinomas (18), experiments demonstrating the malignant transformation of adenoma cell lines (27), and studies identifying common mutations in adenomatous and carcinomatous tissue (28).

Question 2: Intervention/Exposure-SEC

For a given SEC to be valid with respect to a given intervention, we need to demonstrate that the intervention results in a change in the SEC or, in an observational setting, that the exposure of interest is associated with marker positivity.

We can address this question in small clinical (metabolic) studies with the putative SEC as the end point. Examples include studies of fat (29) or alcohol (30) consumption in relation to serum hormone levels. We can also examine this question in a case-control or cohort study of, for example, the relations of reproductive risk factors to HPV infection or breast cancer risk factors to serum estrogen levels. Ecological studies can also provide some limited information on this question. One could, for example, examine the mean proliferative index or degree of epithelial cell DNA hypomethylation in populations with different (average) consumptions of dietary fat (31).

Question 3: Intervention/Exposure-SEC-Cancer

Suppose we have established that: (a) the SEC is causally connected to cancer, but the AP is <1.0 and the route to cancer does not proceed exclusively through the SEC; and (b) the intervention or exposure of interest is linked to the SEC. We would still like to determine the relative importance of the intervention/exposure-SEC-pathway, as opposed to pathways operating through other markers. To do this, we examine the extent to which the relation of exposure/interventions to cancer is mediated by the SEC. That is, we address whether SEC status accounts for the observed intervention effect or exposure-associated elevation in risk. This involves integrating SEC assays into either observational epidemiological studies or clinical trials.

In a recent case-control study, for example, Schiffman and Schatzkin examined the extent to which HPV infection mediated the relation between number of sexual partners and cervical dysplasia (32). As Table 1 indicates, there was a strong direct association between number of sexual partners and risk of cervical dysplasia. When the relation between number of sexual partners and cervical dysplasia was adjusted for presence or absence of HPV infection, the RR for number of sexual partners dropped dramatically, suggesting that most of the association between number of partners and dysplasia is attributable to HPV infection.

Investigators who obtain blood specimens from participants in large cohort studies will be able to investigate whether serum hormone levels mediate the relation between reproduc-

Table 1 Cervical dysplasia odds ratio for number of sexual partners, unadjusted and adjusted for HPV status

	No. of sexual partners				
	1	2	3-5	6-9	10+
Unadjusted	1.0	1.7	3.1 ^a	4.7 ^a	4.4 ^a
Adjusted for HPV status	1.0	1.0	1.1	1.5	1.6

^a*P* < 0.05.

tive risk factors and breast cancer (33). A dietary intervention study of colorectal neoplasia, in which rectal biopsy specimens are obtained for assays of epithelial cell proliferation, has the potential to yield information on the extent to which any dietary effect is attributable to changes in proliferation (34).

One can examine the mediating role of a potential SEC through stratified analyses or standard multiple regression techniques. In general, the larger the intervention effect or exposure relation, the fewer study participants are needed in a mediation analysis. Because exposure RRs in observational epidemiological studies are often greater than the intervention effects observed in trials, mediation analyses are more likely to be successful in the observational epidemiological setting. Genetic mutations as exposures for cancer may prove to be a very fruitful source for mediation analyses of biochemical or cellular markers, if they demonstrate the very high RRs that are currently predicted.

Mediation analyses may yield null results. That is, adjusting for a potential SEC may have little influence on the RRs for the intervention or exposure. These null findings suggest that the potential SEC does not mediate the relation between intervention/exposure and cancer. Even in the face of such null results, however, there are two circumstances under which the SEC could still reside on the causal pathway to cancer. The first, illustrated in Figs. 2 and 7, is when there is an alternative pathway from the exposure (E1) to cancer through a second marker. That is, the SEC is not a necessary step between E1 and cancer. The degree to which the E1-cancer relation is attenuated after adjustment for the SEC will depend on the (likely unknown) extent to which the E1-cancer relation is mediated by MARKER2 as well as the SEC.

The second circumstance is illustrated in Fig. 9. Some unknown factor leads to the SEC. In addition, the SEC requires the exposure E1 as a cofactor for the development of cancer. The SEC is on the pathway to cancer, but adjustment for marker status will not necessarily reduce the RR of the exposure to 1.0. The SEC does not mediate the known risk factor but does mediate the unknown risk factor.

An additional consideration in mediation analyses is the possibility of interaction. In Fig. 3, an intervention influences both the SEC and another intermediate marker (MARKER2). It is at least theoretically possible that the intervention can affect SEC and MARKER2 in counterbalancing ways. In that case, the mediation analysis will reveal a significant interaction between the intervention and the SEC, that is, the cancer rate among SEC-positive individuals will differ according to whether they are in the intervention (exposed) or control (nonexposed) group. Such an interaction indicates that the SEC does not fully mediate the intervention effect (35). However, the SEC does indeed lie on a single dominant causal pathway.

Unknown risk factor → SEC + E1 → CA

Fig. 9. SEC is on the pathway to cancer (CA), but SEC adjustment has little or no influence on the observed association between exposure (E1) and cancer.

Interpreting the Data on SECs: Epidemiological and Statistical Considerations

As in any epidemiological study or clinical trial, one cannot escape the need to apply common sense and judgment to population studies involving SECs. The traditional epidemiological criteria of causality can be useful in evaluating data from these SEC studies. Are the results biologically plausible? Yes, there is good reason to think that HPV infection might explain the strong association between number of sexual partners and cervical neoplasia. Are data from multiple studies consistent? Several studies have shown the relation between HPV and cervical neoplasia, and a few have now demonstrated that HPV mediates the reproductive risk factor-cervical neoplasia association. Are the measures of effect (the RR, AP) strong? They are for HPV-cervical neoplasia. Is the reduction of RR in the mediation analyses substantial? It was in the number of sexual partners-HPV infection-cervical dysplasia study. In other words, in evaluating the validity of a potential SEC, one must examine the totality of evidence.

All markers are measured with some error. Two statistical caveats follow from this. First, a potential SEC is useful only if it can discriminate among study participants, those in an intervention and control group or those in various categories of risk factor exposure. Such discrimination is practically possible only if the inter-participant variation in the SEC values is not swamped by intra-individual variation. (Intra-individual variation derives, for example, from differences in markers obtained from different tissue areas, measured at different time points, or read by multiple readers.) Statistically, this means that the ICC (defined below) for inter-participant variation (the proportion of all variation attributable to between-participant differences) is reasonably large (36).

$$ICC = \frac{\text{inter-participant variation}}{\text{inter-participant variation} + \text{intra-participant variation}}$$

Intra-participant variation may be reduced by taking replicate samples (multiple biopsies from different areas, multiple blood draws over time). The reduction in intra-participant variability increases the relative contribution of the inter-participant variability and, thus, the ICC. At a minimum, therefore, data are required on components of variance to establish the minimum number of marker samples that are needed for meaningful discrimination among study participants. Without such data, one cannot be certain that null marker findings reflect true absence of effect or association or simply the attenuating influence of random sources of intra-individual variation.

These data have been sparse. Few studies have provided data on SEC variability, particularly with respect to time-to-time variation. A notable exception are recent investigations attempting to determine the number of estradiol measurements necessary to reasonably discriminate among individuals (37-39). Research into rectal mucosal proliferation variability is also underway (7, 40). We emphasize that quality control studies designed to capture information on marker variability are essential if we wish to evaluate and subsequently use a potential SEC.

Second, and more generally, even if the ratio of inter- to intra-participant variation is acceptable, measurement error

will tend to attenuate findings from each of the three types of studies discussed above. The intervention-marker and marker-cancer relations will be attenuated by error in marker measurement. Similarly, in mediation analyses, the expected attenuation of the intervention effect will itself be attenuated; in other words, the marker-adjusted intervention effect will be inflated.

Other Issues in Evaluating SECs

A two-stage strategy whereby one examines the relation of an intervention or exposure to a SEC and, separately, the relation of the marker to cancer can provide strong (but not definitive) evidence that the intervention/exposure is truly related to cancer. One could look, for example, at whether alcohol affects estrogen levels and also whether estrogen levels are associated with breast cancer. This two-stage evidence is less than definitive, however, because of the possibility that the intervention/exposure is related to a second marker in a way that offsets the effect through the first marker (a possibility reflected in Figs. 2 and 7).

A marker may not be directly on the causal pathway to cancer but may be closely linked to a component of that causal pathway such that it does make a reasonable SEC. One possible example is micronuclei, which have been detected in epithelial cells from oral, esophageal, bronchial, and large intestinal tissue (41). Many micronucleated cells are nonviable and therefore cannot be a direct cellular precursor of a malignant tumor. The overall prevalence of micronucleated cells, however, might strongly reflect microstructural alterations in other cells that do eventually undergo malignant transformation and clonal expansion.

Comparing one potential SEC to another (for example, proliferating cell nuclear antigen to bromodeoxyuridine or tritiated thymidine cell proliferation assays; Ref. 42) can provide useful information on assay characteristics but does not constitute validation. The close association of a newer marker with an older "gold standard" marker does not in itself surmount inferential limitations of the older marker.

Conclusion

There is no denying the attractiveness, in a tightening fiscal environment, of conducting studies with surrogate end points. Such studies can be suggestive and may give the right answers about the effect of an intervention or the association with an exposure. We acknowledge a legitimate place for some Phase 2 studies that use SECs. Positive results from these studies provide additional (but not incontrovertible) support for moving on to the larger, more expensive, Phase 3 studies with cancer end points.

Merely being on the causal pathway to cancer does not in itself constitute surrogate validity. It is the totality of causal connections that is critical. Only when the causal pathway goes predominantly through that SEC and the intervention does not have offsetting effects on the SEC and a subsequent marker (following Fig. 3) can one reasonably make strong inferences from SEC findings to cancer. This is likely to be the case for adenomas, HPV infection, and Barrett's esophagus.

When major alternative pathways to cancer bypass a potential SEC, inferences to cancer are problematic. This paper is in part a plea to carry out the research (especially SEC-cancer and intervention/exposure-SEC-cancer mediation studies) necessary to evaluate such SECs. These studies are urgently needed if we are to know how well we can generalize from SEC results to cancer. The irony of the surrogate marker problem,

however, is that the large, long, expensive studies required to evaluate SECs fully are the same studies the markers were designed to replace. Moreover, SEC evaluation is often intervention/exposure-dependent; results of validation studies of a marker in relation to one intervention/exposure are not necessarily transferable to the marker in relation to another intervention/exposure. Thus, even if we were to find out from some ongoing calcium intervention studies that cell proliferation is a reasonably valid surrogate for colon cancer, we cannot be certain that proliferation indices are equally valid in aspirin or folate trials.

When it comes to obtaining the hard evidence needed for making public health policy, there seems to be no substitute for carrying out large-scale epidemiological studies and clinical trials with cancer end points or, at a somewhat lower level of inferential certainty, SECs that are (for the most part) obligatory steps on the causal pathway to malignant disease.

References

1. Prentice, R. L. Surrogate end points in clinical trials: definition and operational criteria. *Stat. Med.*, **8**: 431-440, 1989.
2. Schatzkin, A., Freedman, L. S., Dawsey, S. M., and Lanza, E. Interpreting precursor studies: what polyp trials tell us about large bowel cancer. *J. Natl. Cancer Inst.*, **86**: 1053-1057, 1994.
3. Mitchell, M. F., Hittelman, W. N., Hong, W. K., Lotan, R., and Schottenfeld, D. The natural history of cervical intraepithelial neoplasia: an argument for intermediate end point biomarkers. *Cancer Epidemiol., Biomarkers & Prev.*, **3**: 619-626, 1994.
4. Misset, J. L., Mathé, G., Santelli, G., Gouveia, J., Homasson, J. P., Sudve, N. C., and Gaget, H. Regression of bronchial epidermoid metaplasia in heavy smokers with tretinoin treatment. *Cancer Detect. Prev.*, **9**: 167-170, 1986.
5. Correa, P. A human model for gastric carcinogenesis. *Cancer Res.*, **48**: 3554-3560, 1988.
6. Garewal, H. S. Chemoprevention of oral cancer: β -carotene and vitamin E in leukoplakia. *Eur. J. Cancer Prev.*, **3**: 101-107, 1994.
7. Baron, J. A., Wargovich, M. J., Tosteson, T. D., Sandler, R., Haile, R., Summers, R., van Stolk, R., Rothstein, R., and Weiss, J. Epidemiological use of rectal proliferation measures. *Cancer Epidemiol., Biomarkers & Prev.*, **4**: 57-61, 1995.
8. Bedi, A., Pasrich, P. J., Akhtar, A. J., Barber, J. P., Bedi, G. C., Giardiello, F. M., Zehnauer, B. A., Hamilton, S. R., and Jones, R. J. Inhibition of apoptosis during development of colorectal cancer. *Cancer Res.*, **55**: 1811-1816, 1995.
9. Fearon, E. R. Genetic alterations underlying colorectal tumorigenesis. *Cancer Surv.*, **12**: 119-136, 1992.
10. Feinberg, A. P., and Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature (Lond.)*, **301**: 89-91, 1983.
11. Groopman, J. D., Wogan, G. N., Roebuck, B. D., and Kensler, T. W. Molecular biomarkers for aflatoxins and their application to human cancer prevention. *Cancer Res.*, **54** (Suppl.): 1907s-1911s, 1994.
12. Trizna, Z., Schantz, S. P., Lee, J. J., Spitz, M. R., Goepfert, H., Hsu, T. C., and Hong, W. K. *In vitro* protective effects of chemopreventive agents against bleomycin-induced genotoxicity in lymphoblastoid cell lines and peripheral blood lymphocytes of head and neck cancer patients. *Cancer Detect. Prev.*, **17**: 575-583, 1993.
13. Schiffman, M. H. Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. *J. Natl. Cancer Inst.*, **84**: 394-398, 1992.
14. Munoz, N. Is *Helicobacter pylori* a cause of gastric cancer? An appraisal of the seroepidemiological evidence. *Cancer Epidemiol., Biomarkers & Prev.*, **3**: 445-451, 1994.
15. Oza, A. M., and Boyd, N. F. Mammographic parenchymal patterns: a marker of breast cancer risk. *Epidemiol. Rev.*, **15**: 196-208, 1993.
16. Karlan, B. Y. Screening for ovarian cancer: what are the optimal surrogate end points for clinical trials? *J. Cell. Biochem.*, **23** (Suppl.): 227-232, 1995.
17. Dorgan, J. F., Longcope, C., Stephenson, H. E., Jr., Falk, R. T., Miller, R., Franz, C., Kanle, L., Campbell, W. S., Tangrea, J. A., and Schatzkin, A. Relations of prediagnostic serum estrogen and androgen levels to breast cancer risk. *Cancer Epidemiol., Biomarkers & Prev.*, **5**: 533-539, 1996.
18. Hill, M. J. From adenoma to carcinoma of the colorectum. *Recent Results Cancer Res.*, **122**: 71-84, 1991.

19. Muto, T., Bussey, H. J. R., and Morson, B. C. The evolution of cancer of the colon and rectum. *Cancer* (Phila.), *36*: 2251-2270, 1975.
20. Haggitt, R. C. Barrett's esophagus, dysplasia, and adenocarcinoma. *Hum. Pathol.*, *25*: 982-993, 1994.
21. Berenson, M. M., Johnson, T. D., Markowitz, N. R., Buchi, K. N., and Samowitz, W. S. Restoration of squamous mucosa after ablation of Barrett's esophageal epithelium. *Gastroenterology*, *104*: 1686-1691, 1993.
22. Schatzkin, A., Freedman, L. S., Schiffman, M. H., and Dawsey, S. M. The validation of intermediate end points in cancer research. *J. Natl. Cancer Inst.*, *82*: 1746-1752, 1990.
23. Schiffman, M. H., Bauer, H. M., Hoover, R. N., Glass, A. G., Cadell, D. M., Rush, B. B., Scott, D. R., Sherman, M. E., Kurman, R. J., Wacholder, S., Stanton, C. K., and Manos, M. M. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J. Natl. Cancer Inst.*, *85*: 958-964, 1993.
24. Toniolo, P. G., Levitz, M., Zeleniuch-Jacquotte, A., Banerjee, S., Koenig, K. L., Shore, R. E., Strax, P., and Pasternack, B. S. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J. Natl. Cancer Inst.*, *87*: 190-197, 1995.
25. Schatzkin, A., Lanza, E., Freedman, L. S., Tangrea, J., Cooper, M. R., Marshall, J. R., Murphy, P. A., Selby, J. V., Shike, M., Schade, R. R., Burt, R. W., Kikendall, J. W., and Cahill, J. The Polyp Prevention Trial: I. rationale, design, recruitment, and baseline participant characteristics. *Cancer Epidemiol., Biomarkers & Prev.*, *5*: 375-383, 1996.
26. Lipkin, M., Blattner, W. A., Gardner, E. J., Burt, R. W., Lynch, H., Deschner, E., Winawer, S., and Fraumeni, J. F. Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial nonpolyposis colon cancer from [³H]thymidine labeling patterns in colonic epithelial cells. *Cancer Res.*, *44*: 4201-4207, 1984.
27. Williams, A. C., Harper, S. J., and Paraskeva, C. Neoplastic transformation of a human colonic epithelial cell line: *in vitro* evidence for the adenoma to carcinoma sequence. *Cancer Res.*, *50*: 4724-4730, 1990.
28. Cho, K. R., and Vogelstein, B. Genetic alterations in the adenoma-carcinoma sequence. *Cancer* (Phila.), *70*: 1727-1731, 1992.
29. Prentice, R., Thompson, D., Clifford, C., Gorbach, S., Goldin, B., and Byar, D. Dietary fat reduction and plasma estradiol concentration in healthy premenopausal women. *J. Natl. Cancer Inst.*, *82*: 129-134, 1990.
30. Reichman, M. E., Judd, J. T., Longcope, C., Schatzkin, A., Nair, P. P., Campbell, W. S., Clevidence, B. A., and Taylor, P. R. Effects of moderate alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. *J. Natl. Cancer Inst.*, *85*: 722-777, 1993.
31. Lipkin, M., Uehara, K., Winawer, S., Sanchez, A., Bauer, C., Phillips, R., Lynch, H. T., Blattner, W. A., and Fraumeni, J. F. Seventh-Day Adventist vegetarians have a quiescent proliferative activity in colonic mucosa. *Cancer Lett.*, *26*: 139-144, 1985.
32. Schiffman, M. H., and Schatzkin, A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res.*, *54* (Suppl.): 1944s-1947s, 1994.
33. Riboli, E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann. Oncol.*, *3*: 783-791, 1992.
34. Freedman, L. S., and Schatzkin, A. Sample size for studying intermediate end points within intervention trials or observational studies. *Am. J. Epidemiol.*, *136*: 1148-1159, 1992.
35. Freedman, L. S., Graubard, B. I., and Schatzkin, A. Statistical validation of intermediate end points. *Stat. Med.*, *11*: 167-178, 1992.
36. Fleiss, J. L. *The Design and Analysis of Clinical Experiments*, pp. 1-5. New York: John Wiley & Sons, Inc., 1986.
37. Cauley, J. A., Gutai, J. P., Kuller, L. H., and Powell, J. G. Reliability and interrelations among serum sex hormones in postmenopausal women. *Am. J. Epidemiol.*, *133*: 50-57, 1991.
38. Toniolo, P., Koenig, K. L., Pasternack, B. S., Banerjee, S., Rossenberg, C., Shore, R. E., Strax, P., and Levitz, M. Reliability of measurements of total, protein-bound, and unbound estradiol in serum. *Cancer Epidemiol., Biomarkers & Prev.*, *3*: 47-50, 1994.
39. Hankinson, S. E., Manson, J. E., Spiegelman, D., Willett, W. C., Longcope, C., and Speizer, F. E. Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. *Cancer Epidemiol., Biomarkers & Prev.*, *4*: 649-654, 1995.
40. Lyles, C. M., Sandler, R. S., Keku, T. O., Kupper, L. L., Millikan, C., Murray, S. C., Bangdiwala, S. I., and Ulshen, M. H. Reproducibility and variability of the rectal mucosal proliferation index using proliferating cell nuclear immunohistochemistry. *Cancer Epidemiol., Biomarkers & Prev.*, *3*: 597-605, 1994.
41. Garewal, H. S., Ramsey, L., Kaugars, G., and Boyle, J. Clinical experience with the micronucleus assay. *J. Cell. Biochem.*, *17* (Suppl.): 206-212, 1993.
42. Einspahr, J., Alberts, D., Xie, T., Ritchie, J., Earnest, D., Hixson, L., Powell, M., Roe, D., and Grogan, T. Comparison of proliferating cell nuclear antigen *versus* the more standard measures of rectal mucosal proliferation rates in subjects with a history of colorectal cancer and normal age-matched controls. *Cancer Epidemiol., Biomarkers & Prev.*, *4*: 359-366, 1995.