THE PROMISE AND PERIL OF SURROGATE END POINTS IN CANCER RESEARCH

Arthur Schatzkin and Mitchell Gail‡*

Both experimental and observational studies of cancer need to have an end point. Traditionally, in aetiological and prevention studies, that end point has been the incidence of cancer itself, whereas in therapeutic trials, the end point is usually time to cancer recurrence or death. But cancer takes a long time to develop in an individual and is rare in the population. Therefore, aetiological studies and prevention trials must be large and lengthy to be meaningful. Similarly, many therapeutic trials require a long follow-up of large numbers of patients. Surrogate end points — markers of preclinical cancer or of imminent recurrence — are therefore an attractive alternative. But how can we be sure that a study with a surrogate outcome gives us the right answer about the true end point?

INTERVENTION STUDIES
Also known as clinical trials, these are clinical experiments in which the types of treatment and their allocation to study participants are under the control of the investigator. Usually the treatments are randomly allocated to study participants.

**Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892-7232, USA.*
‡Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892-7244, USA.
e-mail: schatzka@mail.nih.gov
DOI: 10.1038/nrc702

Cancer is one of humanity's leading causes of morbidity and mortality. Nevertheless, in the general population, even the most common malignancies have a low probability of occurrence over a restricted time interval. For example, the age-adjusted annual incidence rate of breast cancer among women in the United States is about 100 per 100,000, or 0.1%; the annual colorectal cancer incidence rate among men and women combined is around 50 per 100,000, or only 0.05%.

The medical research implications of this simple fact are straightforward: controlled INTERVENTION STUDIES or PROSPECTIVE OBSERVATIONAL epidemiological investigations that use incident cancer as an end point must be large, lengthy and costly. Such studies must yield many hundreds of cancers to have adequate statistical power to detect a meaningful treatment effect or exposure association. The ongoing Women's Health Initiative, for example, requires several tens of thousands of participants to be followed over nearly a decade to observe sufficient numbers of cancers to detect reasonable reductions in the incidence of breast and colorectal malignancies[1]. Studies with surrogate end points — biomarkers of preclinical carcinogenesis — are attractive because such studies are potentially smaller, shorter and considerably less expensive than their counterparts with cancer end points.

When are surrogates appropriate?

Despite their potential to reduce the size, duration and cost of studies, surrogate end points might not be acceptable because the quality of evidence they provide on treatment effects or exposure associations is lower than that obtained by studying the effects of treatment or exposure on a true cancer end point. For some types of study, the quality of evidence provided by surrogates might be sufficient, whereas for others only the cancer end points will do. For example, true clinical end points, such as time to cancer recurrence or time to death, might be indispensable in randomized Phase III clinical trials that are designed to estimate the clinical effects of a new cancer treatment. Such trials must provide the highest standards of evidence regarding treatment efficacy. Phase II trials, by contrast, are preliminary studies designed to determine whether an agent warrants further study in Phase III trials, so the use of a surrogate end point, such as whether a tumour shrinks following treatment, might be acceptable. The consequences of a false-negative result might be to curtail testing of a potentially valuable treatment; however, a false-positive result would not lead to widespread use of the agent, but only to Phase III testing in which,

Summary

- Intervention studies or prospective observational epidemiological investigations that use incident cancer as an end point are large, lengthy and costly. Similarly, therapeutic trials based on time to recurrence or mortality can require large numbers of patients and a long follow-up.
- Therefore, studies with surrogate end points — biomarkers of preclinical carcinogenesis — are attractive because they are potentially smaller, shorter and considerably less costly than their counterparts with cancer end points.
- Studies based on surrogate end points, however, are inherently less reliable than studies with the 'true' end point (for example, incident cancer, cancer recurrence or mortality). It is important to know when the use of surrogate end points is appropriate and when it is not.
- A key issue is whether the test of an association between an exposure (or treatment) and a surrogate end point will reliably indicate whether there is an association between the exposure (treatment) and cancer. Three statistical conditions are needed to establish this: first, the surrogate end point is associated with cancer; second, the exposure (treatment) is associated with the surrogate end point; and third, the surrogate end point 'mediates' the association between exposure (treatment) and cancer. Causal pathway diagrams are useful in understanding these conditions.
- A second important issue is whether the magnitude of the association between exposure (treatment) and the surrogate end point predicts the magnitude of the association between exposure (treatment) and cancer. A promising approach to this problem relies on the meta-analysis of a series of studies in which exposure (treatment), surrogate end points and cancer are measured concurrently.
- Even a strong surrogate end point, such as colorectal adenomatous polyps, might not yield definitive results for colorectal cancer.
- Nevertheless, there are settings, such as preliminary evaluations of potential therapeutic agents or exploratory investigations of aetiological factors, in which data based on surrogate end points could pave the way for subsequent definitive studies.

presumably, the agent would be found to have no beneficial clinical effect. Similarly, in epidemiological investigations of, for example, the relationship of dietary factors to colorectal or breast cancer, surrogate end points — such as cell PROLIFERATION INDICES or blood hormone concentrations — might provide valuable exploratory information in the evaluation of a new hypothesis, whereas more rigorous testing of that dietary hypothesis might require the use of frank cancer end points.

PROSPECTIVE OBSERVATIONAL STUDIES
Studies of well-defined groups (cohorts) of individuals for whom exposure data are available initially and for whom follow-up procedures are in place to determine if and when subsequent disease end points arise. The exposures and their allocations to cohort members are not controlled by the investigator.

PROLIFERATION INDICES
Measures of the rate of cell turnover or DNA synthesis derived from one of several proliferation bioassays that are currently available.

Identifying surrogate end points for cancer

To define a surrogate end point (*S*), it is necessary first to define the true clinical end point (*T*). In most observational epidemiological studies, *T* is the occurrence of new ('incident') cancer, usually specified as the age or time of cancer diagnosis. In therapeutic clinical trials, *T* is usually taken as the time from treatment to either cancer recurrence or death. Other clinically meaningful measures that influence how a patient feels or functions can also be used as primary end points[2]. Any measurement other than *T* is a potential surrogate measurement. In a preamble to a proposed accelerated approval rule for drugs, the United States Food and Drug Administration defined a surrogate as follows: "A surrogate end point, or 'marker,' is a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions or survives, and is expected to predict the effect of the therapy"[3].

There are a host of biological phenomena — biomarkers of preclinical carcinogenesis — that could potentially serve as cancer surrogates. With the explosion in molecular and cell biology, this list is growing (BOX 1).

Validating surrogate markers

Once we have found a potential surrogate, how do we determine whether it is a good surrogate marker for the true end point? One useful way is based on hypothesis testing for an association with an intervention or exposure, *E*. For a surrogate *S* to be valid for hypothesis testing, the condition '*S* is not associated with *E*' (the 'null hypothesis') must imply that '*T* is not associated with *E*', and *vice versa*[4]. Later, we discuss three conditions that are required to establish this criterion: first, *S* must be associated with *T*; second, *E* must be associated with *S*; and third, *S* 'mediates' the entire effect of *E* on *T* (that is, in statistical terms, *T* is unrelated to *E*, conditional on *S*). If *S* is valid for hypothesis testing, we know that if we reject the null hypothesis that *S* is associated with *E* (that is, we accept that *S* is associated with *E*), we can conclude that *T* is also probably associated with *E*.

Although validity of hypothesis testing based on *S* is desirable, it would be even more useful if we could predict the magnitude of the effect of *E* on *T* from data on the magnitude of the effect of *E* on *S*. Recent proposals for such prediction are based on analysing a series of studies of treatments in a similar class of treatments[5–7] (BOX 2), and 'trial-level validity' (BOX 2) gives an indication of how reliably the magnitude of the effect of *E* on *T* can be predicted. We now turn to some examples that give insight into these criteria for validating a surrogate marker.

The logic of cancer surrogacy

Suppose, in FIG. 1a, *E1* represents an 'exposure' to some environmental or host factor, anything from a chemopreventive agent to a deleterious risk factor. According to this idealized model, a change in *E1* necessarily alters *S*, which, in turn, modifies the true end point — the likelihood of *T*. As we discuss in the next section, a causal pathway such as that depicted in FIG. 1a implies that *S* is valid for hypothesis testing for the particular factor *E1*, but, without further assumptions, does not necessarily imply that *S* will be valid for hypothesis tests for another exposure, *E2*, nor that the magnitudes of the effects of *E1* on *S* can reliably predict the magnitudes of the effects of *E1* on *T* for a series of exposures (trial-level validity, as described in BOX 2).

The scenario in FIG. 1a rarely occurs. Far more realistic are situations reflected in FIG. 1b. Here, *E1* modulates carcinogenesis through two alternative pathways — one through *S*, the other through another marker, *M2*. In fact, there could be several alternative pathways through *M2*, *M3*, *M4* and so on, but to simplify the presentation we refer here to only one alternative pathway through *M2*. *E1* operates through the alternative *M2* pathway, which means that *S* is not a necessary component of carcinogenesis, so we cannot be assured that *S* is a valid surrogate for hypothesis testing in studies of *E1*.

Box 1 | Types of surrogate end points

Alterations in the characteristics of tissues

'Pre-neoplastic' or frankly neoplastic changes are obvious candidates for surrogate end points. Examples include cervical[35], prostatic[36] and endometrial[37] intraepithelial neoplasia; colorectal adenomatous polyps[38]; bronchial metaplasia (a possible pre-neoplastic state for lung cancer)[39]; and dysplastic changes in the oesophagus[40].

Histological changes detected by imaging

Examples include mammographic parenchymal patterns as a surrogate for breast carcinogenesis[41], and ovarian ultrasound abnormalities in ovarian cancer[42].

Cellular phenomena

Surrogates in this category include several assays of epithelial-cell proliferation, including tritiated thymidine or bromodeoxyuridine incorporation into DNA, proliferating cell nuclear antigen (PCNA) and Ki67 (REF. 43). Measures of apoptosis[44] have recently been proposed as potential surrogate end points, as well as the ratio of proliferation to apoptosis. In AIDS research, $CD4^+$ cell counts and human immunodeficiency virus (HIV) viral load have been used as surrogates for clinical end points[45,46].

Molecular markers

A plethora of potential molecular surrogates have been suggested. Examples include specific somatic mutations in cancer-related genes (such as *RAS* or *TP53*), DNA hypo- and hypermethylation of specific genes and gene-expression products (including those measured using microarrays)[47–49]. Chemical DNA adducts can be considered not only as indicators of exposure (which they might well be), but also as markers of a 'downstream' integrated metabolic process — one occurring temporally and developmentally closer to the malignant outcome than the exposure itself[50].

Infection and inflammation

Infectious processes have been implicated in a number of cancers, and these infections could be viewed as surrogate end points. Examples include infections with human papillomavirus (HPV) in cervical carcinogenesis[51], *Helicobacter pylori* in gastric cancer[52] and human T-lymphotropic virus 1 (HTLV1) in adult T-cell leukaemia[53]. Inflammatory cells and cytokines, which contribute to tumour growth, progression and immunosuppression, could serve as surrogate markers[54].

Bioactive substances in blood and tissue

Examples include blood and tissue oestrogens or androgens, oxidation products and antioxidants (again, in both blood and specific tissues), tissue- or cell-type-specific antigens (such as prostate-specific antigen; PSA) and growth factors. For this category of potential surrogates, the marker — blood oestrogen levels[55], for example — might not be found directly in the target tissue, but might still be considered a potential surrogate end point — in this case, for breast cancer.

Cancer prognostic factors

Potential surrogate end points in cancer treatment studies include time to cancer recurrence (when the true end point is survival) and initial tumour shrinkage (instead of true end point, such as time to tumour recurrence or survival).

The reason for this lack of certainty is that *E1* might influence *M2* in a way that offsets its effect on *S*, the final effect on cancer simply being unknown. If *E1*, for example, were to increase *M2* positivity, *E1* could actually end up increasing cancer incidence, while at the same time reducing *S* positivity and giving at least a superficial impression of being anticarcinogenic. (An example from cardiovascular disease is instructive. High-dose diuretics lower blood pressure, but have little effect on cardiovascular disease mortality in hypertensive patients, possibly because diuretics cause hypokalaemia, which increases risk of sudden death[8].) The relationships in FIG. 1b also make trial-level validity less likely than in FIG. 1a, because the magnitude of the effects of *E* on *T* are less likely to be predictable from the effects of *E* on *S* in a series of such studies.

Another important question, discussed in BOX 3, is whether a surrogate that is valid for one intervention (or exposure) is valid for another.

Epithelial hyperproliferation: a case study

How can we apply this logic to potential surrogates? Cell proliferation assays (BOX 1) have been touted as potential surrogates for cancer owing to the dysregulation of cell growth that characterizes malignancy[9]. But are they valid surrogates? FIG. 2 depicts causal events that are potentially involved in the relationship between hyperproliferation and the neoplastic process in the colorectum. If we focus just on the upper portion of this diagram, we see a single pathway going from normal epithelium to hyperproliferative epithelium to

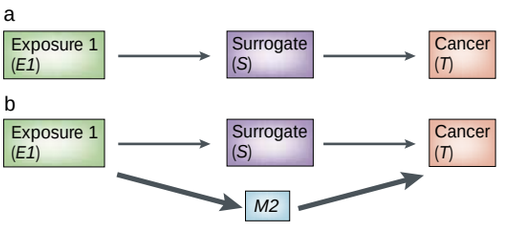
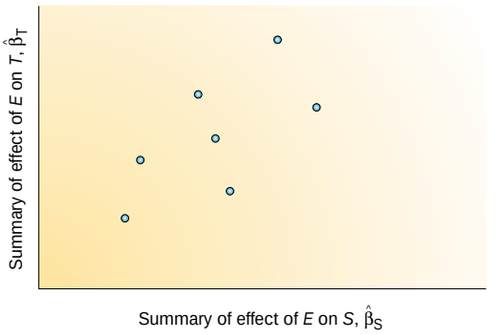


Figure 1 | **Relationships between surrogates and true end points**. **a** | The exposure E1 works through the surrogate marker S to affect the development of cancer (the true end point, T). **b** | E1 affects cancer through two alternative pathways — one through S, the other through a second marker, M2.

Box 2 | Predicting the magnitude of effects on cancer from effects on surrogate end points

Suppose in each study we have sufficient information to allow us to estimate the effect of an exposure, *E* on a surrogate end point, *S* and the effect of *E* on the frank end point, *T*. We might call these two estimated treatment effects or exposure associations $\hat{\beta}_S$ and $\hat{\beta}_T$, obtained by regressing *S* on *E* and *T* on *E*, respectively. In the figure, pairs ($\hat{\beta}_S$, $\hat{\beta}_T$) are plotted for seven different hypothetical clinical trials of various cancer treatments focused on the same molecular pathway, each compared with placebo. If the squared correlation, R^2, among these trial-level pairs was high, we would conclude that the effects of *E* on *S* are highly predictive of the effects of *E* on *T*, and we would say that *S* is 'trial-level valid'[5,6] if the value of R^2 was near 1.0. An analysis of such a series of studies with high R^2 gives us some empirical evidence that if we wish to study a new agent in this same class of agents, we can combine data on the effect of the new agent *E* on *S* with the data from previous studies, as represented in the figure, to predict what the effect of *E* is on *T*. There are, however, a number of limitations to relying on this strategy[7], including potentially serious loss of precision in estimates of the effect of *E* on *T* for the new agent and uncertainty about whether the new agent really belongs to the same class of agents depicted in the figure.



neoplasia/cancer. It is this pathway that implicitly underlies using hyperproliferation as a surrogate for cancer in testing whether there is an association between an exposure and cancer.

But hyperproliferation might not be necessary by itself for colorectal carcinogenesis. There might be an alternative pathway to neoplasia/cancer that bypasses hyperproliferation. The problem is that the effect of an intervention agent (*E1*) on this alternative pathway is unknown and might, in fact, counterbalance the effect through the hyperproliferation pathway. Two scenarios here are revealing: first, the agent (*E1*) reduces proliferation, but at the same time reduces apoptosis, and therefore has no effect on colorectal cancer; second, the agent has no effect on proliferation but does increase apoptosis, thereby reducing colorectal cancer incidence. In both cases, a hyperproliferation assay gives the wrong answer about an intervention's effect on colorectal cancer; by definition, hyperproliferation would not be a valid surrogate for testing for an association between *E1* and cancer.

It is important to emphasize that the proliferation marker does not necessarily give the wrong answer about the agent's effect on cancer; the proliferation data might, in fact, be giving us the right answer. The problem is the uncertainty that flows from the existence of one or more alternative pathways to cancer.

Evaluating potential surrogate end points

Given this uncertainty, how can we evaluate the validity of a potential surrogate marker? The answer is to integrate it into observational epidemiological studies or clinical trials that have cancer (or a preneoplastic lesion, such as adenomatous polyps — see below) as an end point. This integration can elucidate the causal structure underlying the relationships among interventions (or exposures), potential surrogate end points and cancer. In other words, the validation study should include data on *T*, *S* and *E* for each individual. To show a consistent ability to predict the magnitude of the effect of *E* on *T* from data on the effect of *E* on *S* (trial-level validity), there should be a series of such studies.

To determine whether the surrogate is valid for hypothesis testing, we need to find answers to three questions: first, is the potential surrogate associated with cancer incidence (is *S* associated with *T*)?; second, is the exposure or treatment associated with the potential surrogate (is *E* related to *S*)?; and third, does the potential surrogate end point 'mediate' the relationship between exposure or treatment and cancer (that is, conditional on an individual's value of *S*, is there an absence of association between *T* and *E*, as in FIG. 1a)? Standard epidemiological measures, such as RELATIVE RISK and ATTRIBUTABLE PROPORTION, can be used in addressing these questions[10].

Is the surrogate associated with cancer? As indicated above, for a marker to be a reasonable surrogate for a given cancer, it must be associated with that cancer. Ecological studies can provide useful, if indirect, information on this connection. Studies are considered to be 'ecological', or aggregate, when individual-level information is not used; instead, an average marker value is obtained for a sample of individuals selected from specific populations (for example, Seventh Day Adventists versus non-Adventists), which is then related to the overall risk of cancer in those populations. Several studies, for example, have compared mean proliferation indices in groups at varying risk of cancer[11]. In such studies, however, we cannot be certain that those who are marker-positive are the ones with increased incidence of cancer.

This 'ecological' problem is obviated by moving to individual-level observational epidemiological studies, whether case control or cohort. Such studies give individual-level information on *T*, *S* and *E* and they are important tools for examining the relationship between a putative surrogate and cancer. Blood

RELATIVE RISK
An epidemiological measure of treatment effect in an intervention study (clinical trial) or exposure association in a non-experimental observational study. The relative risk is the ratio of risk in an exposed (treated) group to the risk in an unexposed (control) group.

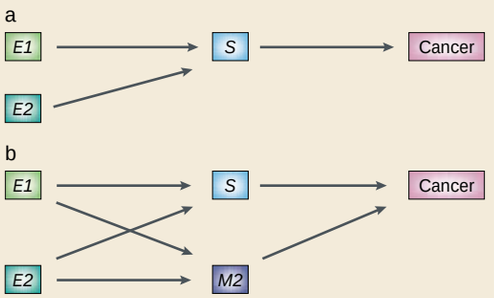
ATTRIBUTABLE PROPORTION (AP). An epidemiological measure of the proportion of all disease cases that is attributable to exposure. The attributable proportion is 1.0 minus the ratio of risk in an unexposed population to the risk in the mixed population of exposed and unexposed individuals. In the context of surrogate markers of cancer, the AP can indicate the proportion of incident cancer that is attributable to marker positivity.

Box 3 | Surrogate validity for different interventions

Is a surrogate that is valid for one intervention valid for another? **a** | Reprises FIG. 1a but adds another exposure, *E2*. (Exposure here can refer to an intervention agent or a risk factor.) Both *E1* and *E2* operate through a single surrogate on the path to cancer. In this scenario, the surrogate is a necessary component of the cancer pathway. *E2* must operate through the surrogate. The surrogate is valid for studies of *E2* as well as those of *E1*. **b** | Here, *E2* enters into the more complex scenario depicted in FIG. 1b. The existence of a non-trivial alternative pathway (through *M2*) means that the validity of the surrogate *S* might be exposure dependent. Even if *E1* works primarily through the surrogate and affects *M2* minimally, indicating that the surrogate is reasonably valid for *E1*–cancer studies, it cannot be assumed that the *E2*–*M2*–cancer pathway has a similarly minor role in carcinogenesis.

For example, a given agent, *E1*, might influence colorectal carcinogenesis largely through its influence on cell proliferation. Cell proliferation in this scenario is likely to be a valid surrogate for colorectal cancer. A second agent, *E2*, might have a minimal effect on cell proliferation 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. The validity of a surrogate must therefore be established for every intervention.

An approach to this problem is to consider studies of a 'class' of biologically comparable intervention agents. If, for example, a meta-analysis shows that the effect of these agents on the surrogate predicts their effect on the true end point, we can be reasonably confident in inferring a treatment effect on the true end point from the effect of a new member of that class on the surrogate end point[5–7], as discussed in BOX 2.



oestrogen levels have been shown in several studies to be directly associated with breast cancer, a relationship that had to be established before oestrogens could be considered a surrogate for breast malignancy[12,13]. Human papillomavirus (HPV) infection, a potential surrogate for cervical cancer, has been shown to be associated with risk of severe cervical neoplasia[14]. Observational studies can also be incorporated into clinical trial design. For example, in the Polyp Prevention Trial[15] — a dietary intervention study with adenomatous polyp formation as the primary end point — investigators are examining the relationship between colorectal epithelial-cell proliferation measures and subsequent adenoma recurrence. (The adenoma or cervical intraepithelial neoplasia (CIN) end points described here are only neoplastic cancer precursors; we have, for purposes of discussion, considered these as proxies for cancer, even though, as we discuss below, the validity of these precursor end points is not ironclad.)

Attributable portion (AP), an epidemiological parameter that measures the extent to which *T* is determined by *S*, can be useful in determining the importance of alternative pathways and thereby evaluating the relationship between *S* and *T*. In the simple linear causal model of FIG. 1a, the estimated AP for the surrogate is 1.0, excluding random error. When at least one pathway exists that is alternative to the pathway containing the surrogate, as in FIG. 1b, then the AP for the surrogate is <1.0. A relatively high AP that was still less than 1.0 would indicate that the alternative ('*M2*') pathway has only a minor function in tumorigenesis. An AP substantially lower than 1.0 for the surrogate implies that one or more alternative pathways is indeed operative.

Is E associated with S? For a potential surrogate marker to be valid with respect to a particular intervention (or exposure), there must be some relationship between the intervention (exposure) and the marker. Ecological studies can provide indirect information on this question. For example, the mean colorectal epithelial-cell proliferation index could be measured in populations with different average consumption of dietary fat. Individual-level studies, however, can provide more convincing evidence.

In a clinical trial, we need to see that the intervention changes the marker, which can be addressed in relatively small studies. Several studies, for example, have examined the effect of dietary change or supplementation on colorectal epithelial-cell proliferation[16]; others have investigated the effect of dietary-fat modification[17] or alcohol consumption[18] (both possible aetiological factors in breast cancer) on blood or urine oestrogen levels. One illustrative case is that no relationship was found between calcium carbonate supplementation and epithelial-cell proliferation measured 1 year later[19], even though calcium did reduce overall adenoma recurrence[20]. This indicates that proliferation measures are problematic surrogates for

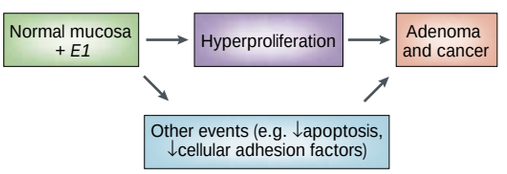


Figure 2 | **Alternative pathways from normal colorectal mucosa to neoplasia (adenoma or cancer).** One pathway involves epithelial hyperproliferation, the other does not.

Table 1 | Number of sexual partners and the risk of cervical dysplasia

	Number of sexual partners				
	1	**2**	**3–5**	**6–9**	**>10**
Odds ratio					
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

*$p < 0.05$. HPV, human papillomavirus.

colorectal neoplasia/cancer in studies with calcium supplements as the main intervention/exposure.

We can also examine this question in case-control or cohort studies, in which we evaluate the association between an exposure and the potential surrogate. Schiffman *et al.*, for example, in investigating the aetiology of cervical cancer, showed a strong association between reproductive risk factors, particularly number of sexual partners, and HPV infection, a potential surrogate for cervical neoplasia[21].

Does S mediate the link between E and T? Once we have determined whether a potential surrogate is highly associated with cancer and whether a surrogate is indeed linked to a given intervention or exposure, it is still necessary to determine whether the effect of *E* on *T* is 'mediated' by *S* to establish the validity of *S* for hypothesis testing. In statistical terms, mediation by *S* means that *E* and *T* are unrelated ('conditionally independent') once marker status is taken into account. One way to test for this condition is to stratify the data on levels of the surrogate marker and determine if there is an association between *E* and *T* within strata. If no such association is present, then there is evidence of mediation. An analogous approach is to include the surrogate marker *S* and the exposure *E* as independent variables in a MULTIPLE REGRESSION model that has *T* as the dependent variable. If the regression coefficient for *E* is 0, this constitutes evidence for mediation. The statistical aspects of mediation analysis are an area of current research[22,23]. Mediation analysis can be misleading if, for example, an intervention has both beneficial and toxic effects and the surrogate captures only the beneficial effects[24].

We can obtain concrete data on mediation by integrating an assay for the surrogate into either clinical trials or observational epidemiological studies, collecting information on both the intervention or exposure and the cancer (or severe neoplasia). As an example, investigators have used a case-control study to look at the extent to which HPV infection mediates the association between number of sexual partners and dysplasia[25]. As TABLE 1 shows, the number of sexual partners was strongly and directly associated with cervical dysplasia risk. When the presence or absence of HPV infection was included as a covariate in a STATISTICAL REGRESSION MODEL that related dysplasia to the number of sexual partners, the relative risk for number of sexual partners dropped dramatically. This indicates that most of the association between number of partners and cervical dysplasia is due to HPV infection[26].

MULTIPLE REGRESSION
A statistical regression model with more than one independent variable.

STATISTICAL REGRESSION MODEL
A statistical approach to quantifying the relationship between an end point ('dependent variable') and other factors ('independent variables') such as treatments or exposures. Regression models are available for continuous, dichotomous, and survival end points.

The same analytical strategy can be used to assess the extent of surrogate mediation in other study designs. For example, researchers obtaining blood specimens from participants in large cohort studies will be able to investigate whether serum hormone levels mediate the relationship between reproductive risk factors and breast cancer. A dietary modification or dietary supplement study of colorectal neoplasia, from which rectal biopsy specimens are obtained for mucosal proliferation assays, could provide information on the extent to which any observed diet/supplement effect is mediated by proliferation changes.

As a general rule, the greater the intervention effect or exposure association, the fewer study participants are needed in a mediation analysis. Because the range of exposure among individual participants in an observational study can be wider than the difference between average treatment-group exposures in intervention studies (trials), the relative risks due to exposures in observational studies tend to be larger than the intervention effects observed in clinical trials. It follows that mediation analyses might be more likely to provide interpretable data in observational epidemiological studies. Although complete mediation is necessary for a marker to be perfectly valid for hypothesis testing, it does not guarantee that the magnitude of the effects of *E* on *S* can be used to predict the magnitude of the effects of *E* on *T* reliably. Moreover, a demonstration that S mediates the effect of *E* on *T* for one exposure does not guarantee that it does so for another exposure. These points highlight the desirability of obtaining data on *E*, *S* and *T* differing exposures (or treatments) (BOX 2).

Surrogates that are likely to be valid

Unlike putative surrogates such as epithelial-cell proliferation or blood hormone levels, for which validity is problematic, considerable evidence supports the usefulness of a few 'downstream' surrogate markers (that is, those close to cancer on the causal pathway).

Cervical cancer surrogates. Practically all cervical cancer requires prior persistent HPV infection. HPV persistence results in inactivation — by the E6 and E7 proteins of the HPV genome — of the *TP53* and *RB* tumour-suppressor genes, leading, in turn, to increasingly severe intraepithelial neoplasia and, eventually, cancer[27]. At most, only a very small proportion of cervical cancer can arise as a result of tumour-suppressor inactivation occurring by mutation in the absence of HPV infection. Because most cervical cancer does occur through persistent HPV infection, an intervention that eliminates or reduces such infection would have a high likelihood of decreasing cervical-cancer incidence.

CIN, especially CIN3, is also considered a strong surrogate for cancer and has been used as an end point in a number of epidemiological studies. A very high percentage of CIN3 will progress to cancer in 20 years; only a very small fraction regresses. In fact, CIN3 is very close to being invasive cancer and is downstream from persistent HPV infection in the causal pathway that leads to malignancy.

Figure 3 | **A limitation of colorectal adenoma recurrence as a surrogate for colorectal cancer.** The intervention in this scenario affects only those 'innocent' adenomas that do not progress to cancer, having no effect on the 'bad' adenomas that lead to malignancy.

Adenomatous polyps for colorectal cancer. Another potential surrogate end point for which inferences to cancer are considered to be strong is the adenomatous polyp (adenoma). Colorectal adenomas are attractive candidates for cancer surrogacy in research studies because of their high recurrence rate: about 10% of persons having an adenoma removed will have a recurrence in the next year — an occurrence frequency nearly two orders of magnitude greater than the incidence of colorectal cancer. The underlying biological rationale for the use of adenoma end points in epidemiological studies and clinical trials is the strong evidence for a relationship between this marker and colorectal cancer (see above section, 'Is the surrogate associated with cancer?'). This adenoma–carcinoma sequence is supported by studies that show carcinomatous foci in adenomas and adenomatous foci within carcinomas, experiments showing the malignant transformation of adenoma cell lines and studies identifying common mutations in adenomatous and carcinomatous tissue[28–30]. An intervention reducing the recurrence of adenomas in the large bowel would therefore probably decrease the incidence of colorectal cancer, thereby making adenoma recurrence a reasonably valid surrogate marker.

Nevertheless, even the adenoma is not a perfectly reliable surrogate and some inferential difficulties remain with trials in which adenoma recurrence is used as a surrogate end point. Recurrent adenomas occur early in the tumorigenic sequence. The results of adenoma recurrence trials can be misleading if the intervention factor being tested operates later in the neoplastic process — for example, from the growth of a small into a large adenoma, or the transformation of a large adenoma to carcinoma. A (false) null result for recurrent adenomas can result if the intervention operates only in the later stages of neoplasia. A positive result, however, indicates that cancer would be reduced, because large adenomas and cancers derive from small adenomas.

A second inferential difficulty with adenoma recurrence as a surrogate end point flows from the likely biological heterogeneity of adenomas. Only a relatively small proportion of adenomas develop into cancer. Suppose that one type, the 'bad' adenoma that progresses to cancer, is caused by exposures *E1* and *E2*, as in FIG. 3. The second type, the 'innocent' adenoma, is caused by the same exposure (*E1*), but only when combined with another exposure, *E3*. Imagine an intervention that works only on *E3*. We could reduce the pool of innocent adenomas — thereby yielding a statistically significant reduction in adenoma formation in our trial — but, in fact, the incidence of bad adenomas and cancer would be unaffected. This could work the other way as well: we might see at most a small reduction in all adenomas (the bad ones being only a small proportion of all adenomas) even though the intervention decreases the formation of bad adenomas and, therefore, reduces the incidence of cancer.

Measurement error

All biomarkers are measured with some error. Two important statistical issues need to be considered. First, a potential surrogate is useful (and ultimately valid) only if it can discriminate among study participants: those in the different treatment arms of a trial or the various exposure categories in an epidemiological study. Discrimination is possible only if the surrogate values vary more between participants than they do within the same individual (due to differences, for example, in marker values obtained from different tissue areas, measured at different time points, or read by several readers.) This can be measured by calculating a value known as the intra-class correlation coefficient (ICC), and this needs to be relatively large if the surrogate is to be useful[31].

Intra-participant variability can be reduced — and the ICC thereby increased — by taking repeat samples, such as several biopsies from different areas or multiple blood samples over time. At a minimum, therefore, data are required on the potential surrogate marker's components of variance to establish the minimum number of marker samples needed for meaningful discrimination among study participants. In the absence of such data, it is not possible to ascertain whether null findings for a potential surrogate reflect a true lack of effect (or association) or simply the attenuating influence of random sources of intra-individual variation.

Reliability data have not been routinely collected in marker studies. Few studies have provided data on potential surrogate-marker variability, particularly with respect to variability over time. A notable exception is recent investigations that attempt to estimate the number of oestradiol measurements necessary to discriminate among individuals[32]. Studies measuring intra-individual variation in colorectal epithelial-cell proliferation are underway[33]. Quality-control studies designed to obtain data on the variability characteristics of potential surrogate markers are essential.

Second, even if the ICC is acceptable, measurement error will tend to attenuate findings from studies that are designed to answer each of the three questions posed above. The associations between intervention (exposure) and marker, and between marker and cancer, will be attenuated by errors in marker measurement. Measurement error in *S* can also lead to an underestimate of the extent to which a correctly measured *S* would mediate the effect of *E* on *T*.

Conclusion

Because studies with surrogate cancer end points can be smaller, faster and substantially less expensive than those with frank cancer outcomes, the use of surrogate end points is undeniably attractive. This attractiveness is likely to grow in the coming years as the rapidly advancing discoveries in cell and molecular biology generate new therapies that require testing, as well as new markers that could plausibly serve as surrogates for cancer.

Surrogate end-point studies can certainly yield useful information. They continue to have a legitimate role in Phase II clinical studies. In some areas of clinical therapeutics, surrogate end points such as blood pressure, blood sugar level or HIV viral load are regarded as useful for Phase III studies. In other circumstances, the most that can be said is that surrogates might give the right answers about intervention effects on (or exposure associations with) cancer.

The problem is the uncertainty attached to conclusions based on surrogates. Except for those few surrogates that are both necessary for and relatively close developmentally to cancer — such as CIN3 and cervical cancer — the existence of plausible alternative pathways makes inferences to cancer from surrogates problematic. Merely being on the causal pathway to cancer does not in itself constitute surrogate validity; it is the totality of causal connections that is crucial. There is, unfortunately, a fairly extensive history of plausible surrogate markers that give the wrong answer about the effects of treatments for chronic disease[34]. There is no reason to believe that observational studies of cancer aetiology based on cancer surrogates are immune to such inferential difficulties.

We should also consider the use of surrogate markers in the broader context of multiple disease end points, including treatment toxicity. A surrogate marker might give the 'right' answer about cancer for a given intervention, but nevertheless give little or no information about important adverse events that greatly influence overall evaluation of the intervention. Suppose, for example, that we have a valid tissue or blood marker for breast cancer — one that gives us the right answer about a promising hormone-modulating intervention. That breast-cancer surrogate will tell us nothing about the potential of the intervention to increase the incidence of stroke. A potential stroke surrogate could be measured, but we are then faced with uncertainties about the reliability of this surrogate for stroke itself. This illustrates yet another difficulty arising from exclusive reliance on surrogate-marker studies.

This article emphasizes the importance of conducting the investigations necessary to evaluate potential surrogates, which include information on *E*, *S* and *T* for study participants. Such studies are needed if we are to generalize from surrogate end-point findings to cancer. There is, however, an implicit and perhaps unavoidable irony here: the large, long, expensive studies required to fully evaluate potential surrogates are precisely the studies that surrogates were designed to replace. Moreover, the exposure dependence alluded to above complicates matters further: establishing validity for a given surrogate for one intervention/exposure does not necessarily translate into validity for another intervention/exposure. To assess validity for a variety of related interventions or exposures, the investigator needs a series of studies that provide individual-level data on *T*, *S* and *E*.

The problems inherent in using surrogate end points need not be regarded as a cause for pessimism in cancer research. If anything, the limitations of surrogacy remind us of the complexity of cancer causation and affirm the continued importance of large clinical trials and observational epidemiological studies with explicit cancer end points. In the context of such a research programme, we might identify surrogates that are useful in exploratory investigations and Phase II trials, and, in some instances, in more definitive studies.

1. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control. Clin. Trials* **19**, 61–109 (1998).
2. DeGruttola, V. G. *et al*. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. *Control. Clin. Trials* (in the press).
   Reviews useful surrogates and Phase III clinical trials, in which their use is often problematic. Sets out statistical concepts and approaches to evaluating surrogate markers, and an agenda for research and resources needed to promote the proper use of surrogates.
3. New drug, antibiotic and biological drug product regulations: accelerated approval. Proposed Rule. **57** *Federal Register* 13234–13232, 1992.
4. Prentice, R. L. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* **8**, 431–440 (1989).
   Defines criteria for the validity of a surrogate marker for testing the hypothesis of no association between a treatment and a true clinical outcome.
5. Daniels, M. J. & Hughes, M. D. Meta-analysis for the evaluation of potential surrogate markers. *Stat. Med.* **16**, 1965–1982 (1997).
6. Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. & Geys, H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 49–67 (2000).
   Defines the concept of trial-level validity of a surrogate, based on meta-analysis of a set of trials. Illustrates the use of meta-analytical data for predicting the size of the effect of a new treatment on the true clinical outcome from data on its effect on the surrogate outcome.
7. Gail, M. H., Pfeiffer, R., Houwelingen, H. C. W. & Carroll, R. J. On meta-analytic assessment of surrogate outcomes. *Biostatistics* **1**, 231–246 (2000).
8. Temple, R. Are surrogate markers adequate to assess cardiovascular disease drugs? *J. Am. Med. Assoc.* **282**, 790–795 (1999).
9. Wargovich, M. J. in *Prevention and Early Detection of Colorectal Cancer* (eds Young, G. P., Rozen, P. & Levin, B) 89–101 (London, W.B. Saunders Co. Ltd, 1996).
10. Rothman, K. J. & Greenland, S. *Modern Epidemiology* (Philadelphia, Lippincott–Raven, 1998).
11. Lipkin, M. *et al*. Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [^{3}H]thymidine labeling patterns in colonic epithelial cells. *Cancer Res.* **44**, 4201–4207 (1984).
12. Toniolo, P. G. *et al*. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J. Natl Cancer Inst*. **87**, 190–197 (1995).
13. Hankinson, S. E. *et al*. Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. *J. Natl Cancer Inst.* **90**, 1292–1299 (1998).
14. Schiffman, M. H. *et al*. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J. Natl Cancer Inst.* **85**, 958–964 (1993).
15. Schatzkin, A. *et al*. Lack of effect of a low-fat, high-fiber, diet on the recurrence of colorectal adenomas. *N. Engl. J. Med.* **342**, 1149–1155 (2000).
16. Holt, P. R. *et al*. Modulation of abnormal colonic epithelial cell proliferation and differentiation by low-fat dairy foods: a randomized controlled trial. *J. Am. Med. Assoc.* **280**, 1074–1079 (1998).
17. Prentice, R. *et al*. Dietary fat reduction and plasma estradiol concentration in healthy premenopausal women. *J. Natl Cancer Inst.* **82**, 129–134 (1990).
18. Reichman, M. E. *et al*. Effects of moderate alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. *J. Natl Cancer Inst.* **85**, 722–727 (1993).
19. Baron, J. A. *et al*. Calcium supplementation and rectal mucosal proliferation: a randomized controlled trial. *J. Natl Cancer Inst.* **87**, 1303–1307 (1995).
20. Baron, J. A. *et al*. Calcium supplements for the prevention of colorectal adenomas. *N. Engl. J. Med.* **340**,101–107 (1999).
21. Schiffman, M. H. *et al*. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J. Natl Cancer Inst.* **85**, 958–964 (1993).

22. Freedman, L. S., Graubard, B. I. & Schatzkin, A. Statistical validation of intermediate endpoints for chronic diseases. *Stat. Med.* **11**, 167–178 (1992).
23. Buyse, M. & Molenberghs, G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* **54**, 1014–1029 (1998).
24. DeGruttola, V., Fleming, T., Lin, D. Y. & Coombs, R. Perpsective: validating surrogate markers — are we being naive? *J. Infect. Dis.* **175**, 237–246 (1997).
25. Schiffman, M. H. & Schatzkin, A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res.* **54**, S1944–S1947 (1994).
26. Franco, E. L. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology* **2**, 98–106 (1991).
27. zur Hausen, H. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J. Natl Cancer Inst.* **92**, 690–698 (2000).
28. Sugarbaker, P. H., Gunderson, L. L. & Wittes, R. E. in *Cancer: Principles and Practice of Oncology* (eds DeVita, V. T. Jr, Hellman, S. & Rosenberg, S. A.) 795–884 (J.B. Lippincott Co., Philadelphia, 1985).
29. Paraskeva, C. *et al*. Colorectal carcinogenesis: sequential steps in the *in vitro* immortalization and transformation of human colonic epithelial cells. *Anticancer Res.* **10**, 1189–1200 (1990).
30. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
31. Fleiss, J. L. *The Design and Analysis of Clinical Experiments* 1–5 (New York, John Wiley & Sons, 1986).
32. Hankinson, S. E. *et al*. Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period. *Cancer Epidemiol. Biomarker Prev.* **4**, 649–654 (1995).
33. Lyles, C. M. *et al*. Reproducibility and variability of the rectal mucosal proliferation index using proliferating cell nuclear antigen immunohistochemistry. *Cancer Epidemiol. Biomarker Prev*. **3**, 597–605 (1994).
34. Fleming, T. R. & DeMets, D. L. Surrogate end points in clinical trials: are we being misled? *Ann. Intern. Med.* **125**, 605–613 (1996).
Presents several examples from clinical studies of surrogate end point findings that did not agree with true end-point results.
35. Mitchell, M. F., Hittelman, W. N., Hong, W. K., Lotan, R. & Schottenfeld, D. The natural history of cervical intraepithelial neoplasia: an argument for intermediate endpoint biomarkers. *Cancer Epidemiol. Biomarker Prev.* **3**, 619–626 (1994).
36. Bostwick, D. G. Prostatic intraepithelial neoplasia is a risk factor for cancer. *Semin. Urol. Oncol.* **17**, 187–198 (1999).
37. Mutter, G. L. Endometrial intraepithelial neoplasia (EIN): will it bring order to chaos? *Gynecol. Oncol.* **76**, 287–290 (2000).
38. Schatzkin, A., Freedman, L. S., Dawsey, S. M. & Lanza, E. Interpreting precursor studies: what polyp trials tell us about large bowel cancer. *J. Natl Cancer Inst.* **86**, 1053–1057 (1994).
39. Misset, J. L. *et al*. Regression of bronchial epidermoid metaplasia in heavy smokers with etretinate treatment. *Cancer Detect. Prev.* **9**, 167–170 (1986).
40. Dawsey, S. M. *et al*. Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. *Cancer* **83**, 220–231 (1998).
41. Saftlas, A. F. *et al*. Mammographic parenchymal patterns as indicators of breast cancer risk. *Am. J. Epidemiol.* **129**, 518–526 (1989).
42. Karlan, B. Y. Screening for ovarian cancer: what are the optimal surrogate endpoints for clinical trials? *J. Cell Biochem.* **23** (Suppl.), 227–232 (1995).
43. Baron, J. A. *et al*. Epidemiological use of rectal proliferation measures. *Cancer Epidemiol. Biomarker Prev.* **4**, 57–61 (1995).
44. Bedi, A. *et al*. Inhibition of apoptosis during development of colorectal cancer. *Cancer Res.* **55**, 1811–1816 (1995).
45. Tsiatis, A. A., DeGruttola, V. & Wulfsohn, M. S. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Am. Stat. Assoc.* **90**, 27–37 (1995).
46. Ruiz, L. *et al*. Plasma HIV-1 RNA as a predictor of the efficacy of adding zalcitabine to a previous regimen with zidovudine. *Antivir. Ther.* **1**, 220–224 (1996).
47. Fearon, E. R. Genetic alterations underlying colorectal tumorigenesis. *Cancer Surveys* **12**, 119–136 (1992).
48. Counts, J. L. & Goodman, J. I. Alterations in DNA methylation may play a variety of roles in carcinogenesis. *Cell* **83**, 13–15 (1995).
49. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, S33–S37 (1999).
50. Groopman, J. D., Wogan, G. N., Roebuck, B. D. & Kensler, T. W. Molecular biomarkers for aflatoxins and their application to human cancer prevention. *Cancer Res.* **54**, S1907–S1911 (1994).
51. Schiffman, M. H. Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. *J. Natl Cancer Inst.* **84**, 394–398 (1992).
52. Munoz, N. Is *Helicobacter pylori* a cause of gastric cancer? An appraisal of the seroepidemiological evidence. *Cancer Epidemiol. Biomarker Prev.* **3**, 445–451 (1994).
53. Blattner, W. A. in *Viral Infections in Humans* 3rd edn (ed. Evans, A. S) 545–592 (Plenum Medical Book Co., New York, 1989).
54. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **357**, 539–545 (2001).
55. Dorgan, J. F. *et al*. Relations of prediagnostic serum estrogen and androgen levels to breast cancer risk. *Cancer Epidemiol. Biomarker Prev.* **5**, 533–539 (1996).

Online links

DATABASES

The following terms in this article are linked online to:

CancerNet: http://cancernet.nci.nih.gov/
breast cancer | cervical cancer | colorectal cancer | endometrial cancer | oesophageal cancer | ovarian cancer | prostatic cancer

GenBank: http://www.ncbi.nlm.nih.gov/
E6 | E7

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/
PCNA | PSA | *RAS* | *RB* | *TP53*

FURTHER INFORMATION

FDA Center for Drug Evaluation and Research — approval of drugs based on surrogate end points:
http://www.fda.gov/cder/rdmt/accapp.htm

NCI Biometric Research Branch: http://linus.nci.nih.gov/~brb/

NCI Division of Cancer Epidemiology and Genetics:
http://www.dceg.cancer.gov/

NCI Early Detection Research Network:
http://edrn.nci.nih.gov/index.html

Tutorials on randomized clinical studies from Beth Israel Deaconess Biometrics Center:
http://research.bidmc.harvard.edu/BIDBC/

Access to this interactive links box is free online.