Surrogate End Points in Clinical Trials: Are We Being Misled?

Thomas R. Fleming, PhD, and David L. DeMets, PhD

Phase 3 clinical trials, which evaluate the effect that new interventions have on the clinical outcomes of particular relevance to the patient (such as death, loss of vision, or other major symptomatic event), often require many participants to be followed for a long time. There has recently been great interest in using surrogate end points, such as tumor shrinkage or changes in cholesterol level, blood pressure, CD4 cell count, or other laboratory measures, to reduce the cost and duration of clinical trials. In theory, for a surrogate end point to be an effective substitute for the clinical outcome, effects of the intervention on the surrogate must reliably predict the overall effect on the clinical outcome. In practice, this requirement frequently fails. Among several explanations for this failure is the possibility that the disease process could affect the clinical outcome through several causal pathways that are not mediated through the surrogate, with the intervention's effect on these pathways differing from its effect on the surrogate. Even more likely, the intervention might also affect the clinical outcome by unintended, unanticipated, and unrecognized mechanisms of action that operate independently of the disease process. We use examples from several disease areas to illustrate how surrogate end points have been misleading about the actual effects that treatments have on the health of patients.

Surrogate end points can be useful in phase 2 screening trials for identifying whether a new intervention is biologically active and for guiding decisions about whether the intervention is promising enough to justify a large definitive trial with clinically meaningful outcomes. In definitive phase 3 trials, except for rare circumstances in which the validity of the surrogate end point has already been rigorously established, the primary end point should be the true clinical outcome.

Ann Intern Med. 1996;125:605-613.

From the University of Washington, Seattle, Washington; and the University of Wisconsin–Madison, Medical School, Madison, Wisconsin. For current author addresses, see end of text.

Clinical trials are the standard scientific method for evaluating a new biological agent, drug, device, or procedure for the prevention or treatment of disease in humans. The phase 3 trial is designed to evaluate a new agent's clinical benefit and possible side effects; as such, it is considered to be the definitive test of the agent's usefulness (1–3). For phase 3 trials, the primary end point should be a clinical event relevant to the patient, that is, the event of which the patient is aware and wants to avoid. Examples are death, loss of vision, symptomatic events of the acquired immunodeficiency syndrome (AIDS), the need for ventilatory support, and other events causing a reduction in quality of life. Trials with these clinical outcomes often have a long duration and are expensive. As a consequence, there has recently been great interest in the development of alternative outcomes, or surrogate end points, to reduce the cost and shorten the duration of phase 3 trials (4–17). As defined by Temple (13),

> a surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.

Examples of surrogate end points are increased CD4 cell counts or decreased viral load measures for trials of therapy for human immunodeficiency virus (HIV) infection or AIDS, suppression of ventricular arrhythmias or reduction in cholesterol level or blood pressure in cardiology trials, and tumor regression in trials of cancer therapy. Surrogate end points are rarely, if ever, adequate substitutes for the definitive clinical outcome in phase 3 trials. We review the basic requirements that the surrogate must meet to be used as the replacement outcome.

Requirements for a Surrogate End Point

A correlate does not a surrogate make. It is a common misconception that if an outcome is a correlate (that is, correlated with the true clinical outcome) it can be used as a valid surrogate end point (that is, a replacement for the true clinical outcome). However, proper justification for such replacement requires that the effect of the intervention on the surrogate end point predicts the effect

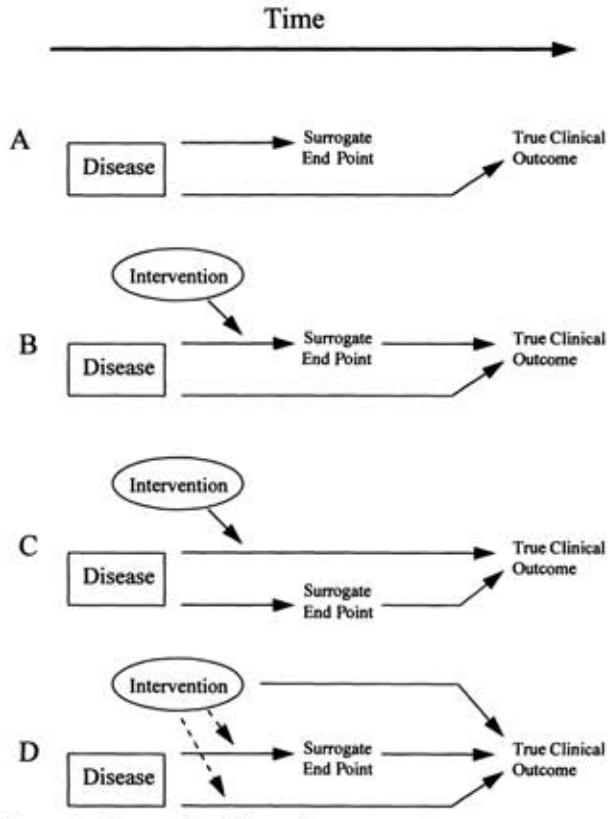

Figure 1. Reasons for failure of surrogate end points. A. The surrogate is not in the causal pathway of the disease process. **B.** Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate. **C.** The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect. **D.** The intervention has mechanisms of action independent of the disease process. *Dotted lines* = mechanisms of action that might exist.

on the clinical outcome—a much stronger condition than correlation.

Prentice (11) developed criteria that are sufficient to validate surrogate end points in phase 3 trials. These criteria essentially require that the surrogate must be a correlate of the true clinical outcome and fully capture the net effect of treatment on the clinical outcome. Although the first criterion is usually easy to verify, the second is not. For example, several recent trials on HIV and AIDS (14–24) showed that the second criterion is not satisfied when CD4 cell count is used as a surrogate end point for development of symptomatic AIDS events or death.

Several factors, illustrated in **Figure 1**, may explain the failure of surrogate end points. Although it may be a correlate of disease progression (**Figure 1A**), a surrogate end point might not involve the same pathophysiologic process that results in the clinical outcome. Even when it does, some disease pathways are probably causally related to the clinical outcome and not related to the surrogate end point. Of the disease pathways affecting the true clinical outcome, the intervention may only affect the pathway mediated through the surrogate end point (**Figure 1B**) or the pathway or pathways independent of the surrogate end point (**Figure 1C**). Most important, the intervention might also affect the true clinical outcome by unintended mechanisms of action that are independent of the disease process (**Figure 1D**). The effects of the intervention mediated through intended mechanisms could be substantially offset by unintended, unanticipated, or unrecognized mechanisms (25).

Figure 2 illustrates the setting that provides the greatest potential for the validity of the surrogate end point. Specifically, the surrogate is in the only causal pathway of the disease process, and the intervention's entire effect on the true clinical outcome is mediated through its effect on the surrogate. Even in this ideal setting, however, surrogate end points can yield misleading conclusions. The intervention's effect on the true clinical end point could be underestimated if there is considerable noise in the measurement of effects on the surrogate end point. The effect on the true end point could be overestimated if the effect on the surrogate, although statistically significant, is not of sufficient size or duration to meaningfully alter the true clinical outcome. This overestimation could readily arise, for example, in the ongoing evaluation of protease inhibitors in HIV-infected patients, in which effects on the surrogate end point (viral RNA levels in the peripheral blood) are substantial but of only short duration.

A review of recent experiences with surrogates is sobering, revealing many cases for which biological markers were correlates of clinical outcomes but failed to predict the effect of treatment on the clinical outcome. In the next section, we examine the failure of surrogates in several clinical trial settings by disease area. We can only speculate about the reasons for these failures because, even in retrospect, our understanding of the causal pathways of the disease process and the mechanisms of action of the intervention is incomplete. **Table 1** provides such speculation, according to the possible explanations provided in **Figure 1**.

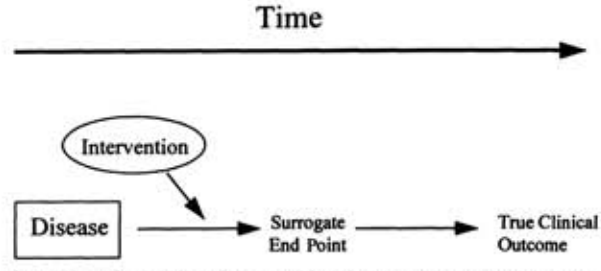

Figure 2. The setting that provides the greatest potential for the surrogate end point to be valid.

Table 1. Speculation on Reasons for Failures of Surrogate End Points*

Disease and Intervention	End Points		Settings in Figure 1†			
	Surrogate	Clinical	A	B	C	D
Cardiologic disorder						
Arrhythmia						
Encainide; flecainide	Ventricular arrhythmias	Survival		+		++
Quinidine; lidocaine	Atrial fibrillation	Survival		+		++
Congestive heart failure						
Milrinone; flosequinan	Cardiac output; ejection fraction	Survival		+		++
Elevated lipid levels						
Fibrates; hormones; diet; lovastatin	Cholesterol levels	Survival		+		++
Elevated blood pressure						
Calcium channel blockers	Blood pressure	Myocardial infarction; survival		+		++
Cancer						
Prevention						
Finasteride	Prostate biopsy	Symptoms; survival	++‡			
Advanced disease						
Fluorouracil plus leucovorin	Tumor shrinkage	Survival		+		++
Other diseases						
HIV infection or AIDS						
Antiretroviral agents	CD4 levels; viral load	AIDS events; survival		+	+	+
Osteoporosis						
Sodium fluoride	Bone mineral density	Bone fractures	+			+
Chronic granulomatous disease						
Interferon-γ	Bacterial killing; superoxide production	Serious infection			++	

* AIDS = acquired immunodeficiency syndrome; HIV = human immunodeficiency virus; + = likely or plausible; ++ = very likely.
† A = surrogate end point not in causal pathway of the disease process; B = of several causal pathways of the disease, the intervention only affects the pathway mediated through the surrogate; C = the surrogate is not in the pathway of the intervention's effect or is insensitive to its effect; D = the intervention has mechanisms of action that are independent of the disease process.
‡ In settings in which only latent disease is prevented.

Surrogate End Points in Cardiology

Arrhythmia Suppression

Use of reduction in ventricular ectopic contractions as a surrogate for decreased cardiovascular-related mortality provides a classic example of the unreliability of surrogate end points. Ventricular arrhythmia is associated with an almost fourfold increase in the risk for death related to cardiac complications (26, 27), particularly sudden death. It was hypothesized that suppression of ventricular arrhythmias after myocardial infarction would reduce the rate of death. Three new drugs (encainide, flecainide, and moricizine) were found to suppress arrhythmias effectively and were approved by the Food and Drug Administration (FDA) for use in patients with life-threatening or severely symptomatic arrhythmias. Although follow-up trials had not been done to determine whether the reduction in arrhythmias would lead to a reduction in death rates, more than 200 000 persons per year eventually took these drugs in the United States. The Cardiac Arrhythmia Suppression Trial (CAST) (26–28) evaluated how the three drugs would affect survival of patients who had had myocardial infarction and had at least 10 premature ventricular beats per hour. The early results from CAST were startling. The encainide and flecainide arms of the trial were terminated early when 33 sudden deaths occurred in patients taking either drug compared with only 9 in the matching placebo group. A total of 56 patients in the encainide and flecainide group died, and 22 patients in the placebo group died. After the data were finalized, the sudden death comparison was 43 and 16 and the number of deaths was 63 in the encainide and flecainide group and 26 in the placebo group. Later results from CAST also established an increased risk for death in patients receiving moricizine (28).

Two other examples are relevant to the arrhythmia setting. Quinidine has been used to maintain sinus rhythm after patients with atrial fibrillation have been converted (29). A meta-analysis of six trials indicated that quinidine maintained sinus rhythm at 1 year (50% of patients who received quinidine compared with 25% of those who did not) but increased the mortality rate from 0.8% to 2.9%. Preventing recurrence of atrial fibrillation is an important benefit, but it does not outweigh the increased mortality rate. Similar inconsistencies were found for lidocaine; a meta-analysis showed that a one-third reduction in the risk for ventricular tachycardia was accompanied by a one-third increase in death rate (30, 31).

Exercise Tolerance in Congestive Heart Failure

Patients with congestive heart failure have decreased cardiac output, characteristic symptoms of dyspnea and orthopnea, decreased exercise capacity, and a high risk for death. The annual mortality rate

for patients with severe congestive heart failure is 20% to 40%. The poor exercise performance is presumed to be a result of decreased cardiac output, but it could also result from increased pulmonary vascular pressure. In this disease, cardiac output and ejection fraction have been used as surrogate end points for examining the usefulness of new drugs, and exercise tolerance and symptomatic improvement have also been regularly assessed as intermediate end points. Although some treatments that affect these end points produce improved survival (32–35), others provide no benefit or actually decrease survival.

Diuretics and digoxin help alleviate symptoms. No data on the survival effects of these treatments have yet been published, although results of the recently reported Digitalis Investigation Group trial (36) show no survival benefit (American College of Cardiology, March 1996. Unpublished data).

One of the earlier drugs that was proposed as a treatment for congestive heart failure was milrinone. Completed studies indicated that milrinone improved cardiac output and increased exercise tolerance. This drug is an inotropic agent (as is digoxin) that stimulates the force of contraction of the heart. Because the FDA was concerned that such agents may have adverse long-term effects (as was the case for β-agonist inotropic agents), a randomized, double-blind, placebo-controlled trial was done to assess mortality rates. The trial, known as PROMISE (Prospective Milrinone Survival Evaluation), showed an increase in total mortality for patients receiving milrinone compared with patients receiving placebo (29% compared with 23%; $P = 0.04$) (37).

Another drug developed for congestive heart failure, flosequinan, is a vasodilator that reduces cardiac workload. This drug was conditionally approved by the FDA because it could improve exercise tolerance in patients who did not respond to or could not tolerate a full regimen of other agents, including diuretics and angiotensin-converting enzyme inhibitors. The conditional approval required completion of a trial, such as the ongoing Prospective Flosequinan Longevity Evaluation (PROFILE) (38), that could evaluate the effect of flosequinan on total mortality. The PROFILE study eventually provided significant evidence that flosequinan increased total mortality (relative risk, 1.43), leading the manufacturer to withdraw the product from the market.

Although cardiac output, ejection fraction, and exercise tolerance are correlated with longer survival of patients with congestive heart failure, a treatment-induced improvement in those measurements is not a reliable predictor of the effect of treatment on mortality rates. Of course, improved exercise tolerance is in itself a clinically relevant outcome and might be considered favorable, despite a potentially adverse effect on mortality. It seems hard to contend, however, that physicians and patients would be able to decide on therapy without knowing the effect on symptoms and on mortality. The effect on mortality can be assessed only through long-term trials.

Lipid Lowering

Although lipid levels, especially those of total cholesterol or its subfractions and triglycerides, have long been known to be significant predictors of cardiovascular-related mortality, researchers have debated the relation between lipid lowering and reduction in overall mortality (39–41). As early as the Coronary Drug Project (CDP) in the 1970s, such drugs as clofibrate and niacin were known to decrease cholesterol levels. However, neither agent reduced total mortality in the highly powered 7-year CDP trial (42).

Many large meta-analyses have been done to evaluate the effects of several types of cholesterol-lowering agents on cause-specific and overall mortality (43–47). Although these analyses differ somewhat in type of treatments and trials that are included, their conclusions are generally consistent with those of Gordon (47), who did the most recent and largest of the meta-analyses. Gordon considered 50 randomized controlled trials of cholesterol-lowering interventions, including diet, fibrates, hormones, resins, and lovastatin. The average reduction in cholesterol level achieved in these trials was 10%. In turn, the intended beneficial effect of reducing the rate of death from coronary heart disease was achieved, with an average reduction of 9%. Unfortunately, these cholesterol-lowering treatments as a group unintentionally increased the mortality rates associated with causes other than coronary heart disease by 24%. In these 50 trials, use of cholesterol-lowering agents actually led to a net 1% increase in overall mortality. Thus, the harmful effects on mortality not related to coronary heart disease completely offset the intended beneficial effects.

A trial in Scandinavian patients with angina pectoris or previous myocardial infarction recently showed that simvastatin decreased cholesterol levels by 25%, with a corresponding 30% reduction in total mortality (48). This is the first major study of a lipid-lowering drug to show a strong overall benefit on mortality. Regardless of whether the reduction in mortality achieved through use of simvastatin has a casual relation with cholesterol lowering, evaluating a treatment only on the basis of its ability to decrease cholesterol levels is clearly inadequate. Without clinical end points, such as total mortality, such drugs as fibrates and hormones could be in widespread use for their cholesterol-lowering effects.

Blood Pressure Lowering

Epidemiologic evidence establishes hypertension as another risk factor for cardiovascular-related mortality (49); a 5% reduction in cardiovascular-related mortality and a 10% reduction in stroke is obtained for every 1 mm Hg reduction in blood pressure. One of the early large studies of treatment for hypertension, the Hypertension Detection and Follow-up Program (50), showed a 17% reduction in total mortality in patients with mild hypertension who were managed with a stepped treatment program beginning with use of diuretics. The more recent Systolic Hypertension in the Elderly Program (SHEP) trial (51) confirmed the benefits on rates of survival, nonfatal myocardial infarction, and stroke provided by a program designed to reduce blood pressure through use of low-dose diuretics.

Current practice in the treatment of hypertension is interesting (52). In many countries, drug approval may be obtained by showing surrogate efficacy; that is, such drugs reduce blood pressure. Since the 1980s, two new classes of drugs, angiotensin-converting enzyme inhibitors and calcium channel blockers, have accounted for nearly 50% of current therapy because of their effect on this surrogate end point and their perceived better side-effect profile compared with those for diuretics and β-blockers. These drugs are as much as 30 times more expensive than diuretics. However, no completed randomized trials have evaluated whether either class of drug reduces the risk for cardiovascular-related mortality or morbidity for hypertensive patients in the primary prevention setting. A recent population-based case–control study suggested that calcium channel blockers may be associated with an increased risk for myocardial infarction among hypertensive patients (53). These drugs have been evaluated in randomized trials in patients with congestive heart failure or myocardial infarction (that is, in the setting of secondary prevention). A recent meta-analysis of survival effects of calcium channel blockers in these trials showed possible harmful effects of this treatment (54). It has been suggested (52) that use of these clinically unproven drugs for the treatment of hypertension may cost an extra $2.5 billion annually in the United States alone.

Although a treatment's effect on blood pressure appears to be a reliable surrogate end point for the evaluation of low-dose diuretics as used in the SHEP trial, this surrogate could be misleading if used to evaluate a new antihypertensive drug. The favorable antihypertensive effects of such agents as calcium channel blockers may be offset by other mechanisms of action that are unanticipated and unrecognized.

Cancer Research

Prevention Trials

Cancer prevention trials attempt to find behavioral modifications or interventions that reduce the risk for cancer in persons at high risk. Because persons who are at high risk and are otherwise healthy may be exposed to the prevention strategy for many years, the need to determine the long-term risk and benefit profile of the prevention strategy is a critical issue.

An important example is a current trial testing finasteride (Proscar, Merck & Co., Inc., West Point, Pennsylvania) for chemoprophylaxis of prostate cancer. If the final clinical outcome of elimination of symptomatic disease or reduction in mortality were used in the finasteride trial, the sample size required to detect prevention effects could be more than 50 000 men. In this trial, a surrogate end point is presence of prostate cancer shown by biopsy after 7 years of follow-up or earlier if clinically indicated. Use of this surrogate reduced the sample size by threefold. Some major concerns with this surrogate are that effects induced by finasteride on the widely used prostate-specific antigen marker will alter the pattern of biopsy sampling, that an estimated 40% of participants will never have biopsy at 7 years, and that finasteride will reduce the volume and alter the texture of the prostate in ways that could differentially affect the rate of false-positive results in the finasteride group and control group. In addition, although approximately 30% of men older than 50 years of age have subclinical prostate cancer (15), only 9% will develop clinical disease and less than 3% will develop fatal disease. Thus, finasteride could reduce the incidence of positive biopsy results and still have no effect on mortality or symptomatic disease. Unless this trial is enlarged or done for a longer period, it is possible that a reduction in prostate cancer proven by biopsy could lead to widespread use of finasteride even though the only tangible effect of the drug could be to harm libido and cause impotence (15).

Treatment Trials

Tumor response has frequently been used as a surrogate end point in therapeutic trials of advanced cancer, especially those that study breast cancer, colorectal cancer, and solid tumors of the lung. The categories of this surrogate end point are complete response (tumor not visible on examination), partial response (a reduction in tumor volume of 50% or more), and no change or progression. Unfortunately, tumor response is not a reliable replacement outcome for survival (3, 55). Many of the trials that have established treatment effects on this

surrogate end point have not shown any change in mortality rates.

The use of surrogates recently produced misleading results in the setting of advanced colorectal cancer. The frequently used treatment of 5-fluorouracil in combination with leucovorin showed a statistically significant improvement in the complete response plus partial response rate (23%) compared with the improvement seen with 5-fluorouracil alone (11%). Despite this difference in tumor response, there was almost no difference in overall survival (relative risk, 0.97). These results were taken from a meta-analysis of almost 1400 patients (56).

Some of the factors contributing to the failure of the surrogate end point (complete response plus partial response rate) are the low proportion of complete responses rather than just partial responses, the proportion of responses that are truly durable long-term effects, and the high likelihood that unintended mechanisms of action from these aggressive and toxic cancer therapies adversely affect survival.

Other Diseases

HIV Infection and AIDS

The use of surrogate end points has probably been more intensely discussed in the design and analysis of clinical trials of HIV infection and AIDS than in any other area. In a review of AIDS trials, Fleming (15) summarized results from the largest trials that evaluated effects of nucleoside analogues on surrogate end points and clinical outcomes. The summary of results from a 1993 state-of-the-art conference (57) shows that the effect of treatment on the most popular surrogate, the CD4 cell count, did not accurately predict the effect of treatment on the clinical outcomes, that is, progression to AIDS or time to death.

In this review, which involved 16 major AIDS trials, the surrogate end point of CD4 cell count was significantly favorable in 7 of the 8 trials in which treatment improved the clinical outcome of progression to AIDS or death. Unfortunately, the CD4 cell count was significantly favorable in 6 of the 8 trials in which treatment did not improve progression to AIDS or death. For survival, the CD4 cell count was significantly favorable in only 2 of 4 trials in which treatment showed a significantly favorable effect on survival and, even worse, was significantly favorable in 6 of 7 trials in which treatment had no effect on survival. Three additional trials, including the Concorde Trial (18), showed an inverse relation between survival and improved CD4 cell counts.

The Concorde Trial (18) involved 1749 asymptomatic HIV-positive patients who were randomly assigned to receive immediate or deferred treatment (when symptoms occurred) with zidovudine. During a follow-up period of 3 years, the decline in CD4 cell counts was slowed by immediate zidovudine therapy, with an average difference of 30 to 35 cells/mm^3 between the two treatment groups. In addition, patients in the group that received deferred treatment with zidovudine more quickly achieved a 50% decline in CD4 cell counts. However, the clinical outcomes did not reflect these changes in the surrogate end point. Time of progression to AIDS-related complex, AIDS, or death was essentially unaffected (175 events in the immediate zidovudine treatment group compared with 171 in the delayed zidovudine treatment group). For death alone, the results actually favored the delayed zidovudine treatment group (95 compared with 76 deaths). The early pressures to use zidovudine treatment in asymptomatic persons with HIV were not supported by these longer-term clinical events.

Osteoporosis in Postmenopausal Women

Postmenopausal women have loss of bone mass and develop osteoporosis, which ultimately leads to an increased risk for fractures of the hip and other bones (58, 59). One strategy is to use therapies, such as estrogen or calcium, to increase bone mass and reduce the incidence of fracture. Sodium fluoride, which stimulates bone formation and increases bone mass, came into widespread use although it was not approved by the FDA. Riggs and colleagues (58) conducted a placebo-controlled randomized trial of fluoride in 202 postmenopausal women who had osteoporosis and vertebral fractures. Patients were followed for 4 years. Treatment increased bone mineral density in the lumbar spine by 35% ($P < 0.001$). However, new vertebral fractures occurred more frequently in patients treated with fluoride than in those who received placebo (163 compared with 136 fractures), and nonvertebral fractures also occurred more frequently in patients treated with fluoride (72 compared with 24 fractures; $P = 0.01$). Riggs and colleagues concluded that the form of fluoride treatment used in their study increased some aspects of bone mineral density but caused bones to become brittle, thereby increasing skeletal fragility.

Chronic Granulomatous Disease

Reliance on surrogate end points also provides a risk for false-negative conclusions that could result in discarding effective treatments. This is shown in a recent trial of chronic granulomatous disease in children (60). Children with this disorder have a compromised immune system: Macrophages engulf

microorganisms but, because they do not generate an oxygen burst, do not kill the microorganisms. As a result, recurrent, serious, and often life-threatening infections develop. Interferon-γ was considered to have therapeutic potential because of its anticipated ability to increase superoxide production and kill bacteria. The initial design of a placebo-controlled trial specified that patients who were randomly assigned to the control group would receive placebo for an interval so brief that only an evaluation of the effect of interferon-γ on the surrogate end points (superoxide production and ability to kill bacteria) would be possible. Before its initiation, the trial was redesigned to enable longer-term assessment of treatment effects on the true clinical outcome, that is, the rate of serious infections. Results of this trial (60), which was conducted by the International CGD Study Group, showed that interferon-γ produced a significant 70% reduction in the rate of recurrent serious infections. It is surprising, however, that this therapy had no detectable effect on the surrogate end points. A shorter-term trial that would have evaluated the effects of interferon-γ only on superoxide production and ability to kill bacteria would have provided a rapid yet unreliable treatment evaluation and would have failed to identify a truly effective treatment.

Conclusions

Effects on surrogate end points often do not predict the true clinical effects of interventions. Although there are many explanations for this failure, such as the existence of causal pathways of the disease process that are not mediated through the surrogate end point and that might be influenced differently by the intervention, the most plausible explanation is usually that the intervention has unintended mechanisms of action that are independent of the disease process. These unintended mechanisms can readily cause the effect on the true clinical outcome to be inconsistent with what would have been expected solely on the basis of evaluation of surrogate end points. These mechanisms are insidious because they are often unanticipated and unrecognized.

Unfortunately, the failure of surrogate end points to predict true outcome is not an isolated problem. **Table 1** shows various examples from several diseases and treatment and prevention strategies, including many comprehensive meta-analyses that involve scores of clinical trials. Several other examples of the failure of surrogate end points can be seen in other settings, ranging from trials of vaccines that use the presence of neutralizing antibodies or cell-mediated immune response as the surrogate end point (61), thrombolytic agent trials that use vessel reperfusion (62–68), and cancer screening strategies that use stage of detected disease (69, 70) to trials of vitamin supplementation for treatment of retinitis pigmentosa using decline of electroretinograms as the surrogate end point (71, 72), oxygen supplementation in severe chronic obstructive pulmonary disease using physiologic variables (73), dental treatments using probing attachment levels (74, 75), and surgery using excision of disease or establishment of blood flow (76).

The validity of a surrogate end point has rarely been rigorously established. Occurrence of false-positive and false-negative results must be low, typically in the range of 2.5% to 10%, in definitive trials evaluating the effects of interventions on clinical outcomes. Hence, to be a valid replacement end point, a surrogate must provide a high level of accuracy in predicting the intervention's effect on the true clinical end point. Predictions having an accuracy of approximately 50%, such as the accuracy seen with the CD4 count in the HIV setting, are as uninformative as a toss of a coin. Methods for validating surrogate end points have been discussed by Lin and colleagues (77), Freedman and associates (78), and DeGruttola and colleagues (79). Statistical methods for validation usually require meta-analyses because the sample sizes needed are much larger than those required for the typical phase 3 evaluation of interventions. Proper validation of surrogates also requires an in-depth understanding of the causal pathways of the disease process as well as the intervention's intended and unintended mechanisms of action. Such insights are rarely achievable.

Surrogate end points should be used where they perform best—in screening for promising new therapies through evaluation of biological activity in preliminary phase 2 trials. Such results in turn can guide decisions about whether the intervention is sufficiently promising to justify the conduct of large-scale and longer-term clinical trials. Although information on surrogate end points in these definitive phase 3 trials can provide further valuable insight into the intervention's mechanisms of action, the primary goal should be to obtain direct evidence about the intervention's effect on safety measures and true clinical outcomes.

Acknowledgments: The authors thank Robert Temple, Curt Furberg, and the reviewers for providing valuable suggestions.

Grant Support: By research grants AI 29168 and CA 18332 from the National Institutes of Health.

Requests for Reprints: Thomas R. Fleming, PhD, Department of Biostatistics, Box 357232, University of Washington, Seattle, WA 98195-7232.

Current Author Addresses: Dr. Fleming: Department of Biostatistics, Box 357232, University of Washington, Seattle, WA 98195-7232.

Dr. DeMets: Department of Biostatistics, K6/446 Clinical Science Center, University of Wisconsin–Madison, Medical School, Madison, WI 53792.

References

1. **Fleming TR.** Evaluating therapeutic interventions: some issues and experiences (with discussion and rejoinder). Statistical Science. 1992;7:428-56.
2. **Fleming TR.** Evaluation of active control trials in AIDS. J Acquir Immune Defic Syndr. 1990;3(Suppl 2):S82-7.
3. **Johnson JR, Temple R.** Food and Drug Administration requirements for approval of new anticancer drugs. Cancer Treat Rep. 1985;69:1155-9.
4. **Ellenberg SS, Hamilton JM.** Surrogate endpoints in clinical trials: cancer. Stat Med. 1989;8:405-13.
5. **Fleming TR, Prentice RL, Pepe MS, Glidden D.** Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. Stat Med. 1994;13:955-68.
6. **Herson J.** The use of surrogate endpoints in clinical trials. Stat Med. 1989; 8:403-4.
7. **Hillis A, Seigel D.** Surrogate endpoints in clinical trials: ophthalmologic disorders. Stat Med. 1989;8:427-30.
8. **Kosorok MR, Fleming TR.** Using surrogate failure time data to increase cost effectiveness in clinical trials. Biometrika. 1993;80:823-33.
9. **Machado SG, Gail MH, Ellenberg SS.** On the use of laboratory markers as surrogates for clinical endpoints in the evaluation of treatment for HIV infection. J Acquir Immune Defic Syndr. 1990;3:1065-73.
10. **Pepe MS, Reilly M, Fleming TR.** Auxiliary outcome data and the mean score method. Journal of Statistical Planning and Inference. 1994;42:137-60.
11. **Prentice RL.** Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989;8:431-40.
12. **Wittes J, Lakatos E, Probstfield J.** Surrogate endpoints in clinical trials: cardiovascular diseases. Stat Med. 1989;8:415-25.
13. **Temple RJ.** A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, eds. Clinical Measurement in Drug Evaluation. New York: J Wiley; 1995.
14. **Ellenberg SS.** Surrogate end points in clinical trials [Editorial]. BMJ. 1991; 302:63-4.
15. **Fleming TR.** Surrogate markers in AIDS and cancer trials. Stat Med. 1994; 13:1423-35.
16. **Lagakos SW, Hoth DF.** Surrogate markers in AIDS: where are we? Ann Intern Med. 1992;116:599-601.
17. **Lin DY, Fischl MA, Schoenfeld DA.** Evaluating the role of CD4-lymphocyte counts as surrogate endpoints in human immunodeficiency virus clinical trials. Stat Med. 1993;12:835-42.
18. **Aboulker JP, Swart AM:** Preliminary analysis of the Concorde Trial. Concorde Coordinating Committee [Letter]. Lancet. 1993;341:889-90.
19. **Choi S, Lagakos SW, Schooley TT, Volberding PA.** $CD4^+$ lymphocytes are an incomplete surrogate marker for clinical progression in persons with asymptomatic HIV infection taking zidovudine. Ann Intern Med. 1993;118: 674-80.
20. **De Gruttola V, Wulfsohn M, Fischl M, Tsiatis A.** Modeling the relationship between survival and CD4 lymphocytes in patients with AIDS and AIDS-related complex. J Acquir Immune Defic Syndr. 1993;6:359-65.
21. **Jacobson MA, Bacchetti P, Kolokathis A, Chaisson RE, Szabo S, Polsky B, et al.** Surrogate markers for survival in patients with AIDS and AIDS related complex treated with zidovudine. BMJ. 1991;302:73-8.
22. **Nowak R.** In Concorde's wake: is the AIDS clinical-trials program flawed? Journal of NIH Research. 1993;5:37-9.
23. **Volberding PA, Lagakos SW, Grimes J, Stein DS, Balfour HH, Reichman RC, et al.** The duration of zidovudine benefit in persons with asymptomatic HIV infection. Prolonged evaluation of protocol 019 of the AIDS Clinical Trials Group. JAMA. 1994;272:437-42.
24. **Volberding PA, Lagakos SW, Koch MA, Pettinelli C, Myers MW, Booth DK, et al.** Zidovudine in asymptomatic human immunodeficiency virus infection. A controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. The AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. N Engl J Med. 1990;322:941-9.
25. **Boissel JP, Collet JP, Moleur P, Haugh M.** Surrogate endpoints: a basis for a rationale approach. Eur J Clin Pharmacol. 1992;43:235-44.
26. Preliminary Report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. N Engl J Med. 1989; 321:406-12.
27. **Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al.** Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. N Engl J Med. 1991; 324:781-8.
28. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. The Cardiac Arrhythmia Suppression Trial II Investigators. N Engl J Med. 1992;327:227-33.
29. **Coplen SE, Antman EM, Berlin JA, Hewitt P, Chalmers TC.** Efficacy and safety of quinidine therapy for maintenance of sinus rhythm after cardioversion. A meta-analysis of randomized control trials. Circulation. 1990;82:1106-16.
30. **Hine LK, Laird N, Hewitt P, Chalmers TC.** Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. Arch Intern Med. 1989;149:2694-8.
31. **MacMahon S, Collins R, Peto R, Koster RW, Yusuf S.** Effects of prophylactic lidocaine in suspected acute myocardial infarction. An overview of results from the randomized, controlled trials. JAMA. 1988;260:1910-6.
32. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). The CONSENSUS Trial Study Group. N Engl J Med. 1987;316:1429-35.
33. **Pfeffer MA, Braunwald E, Moye LA, Basta L, Brown EJ Jr, Cuddy TE, et al.** Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction. Results of the Survival and Ventricular Enlargement Trial. The SAVE Investigators. N Engl J Med. 1992; 327:669-77.
34. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). The CONSENSUS Trial Study Group. N Engl J Med. 1987;316:1429-35.
35. **Feldman AM, Bristow MR, Parmley WW, Carson PE, Pepine CJ, Gilbert EM, et al.** Effects of vesnarinone on morbidity and mortality in patients with heart failure. Vesnarinone Study Group. N Engl J Med. 1993;329:149-55.
36. Rationale, design, implementation, and baseline characteristics of patients in the DIG trial: a large, simple, long-term trial to evaluate the effect of digitalis on mortality in heart failure. The Digitalis Investigation Group. Control Clin Trials. 1996;17:77-97.
37. **Packer M, Carver JR, Rodehoffer RJ, Ivanhoe RJ, DiBianco R, Zeldis SM, et al.** Effect of oral milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. N Engl J Med. 1991;325:1468-75.
38. **Packer M, Rouleau J, Swedberg K, Pitt B, Fisher L, Klepper M, et al.** Effect of Flosequinan on survival in chronic heart failure: preliminary results of the PROFILE study [Abstract]. Circulation. 1993;88(Suppl I):I-301.
39. **Law MR, Thompson SG, Wald NJ.** Assessing possible hazards of reducing serum cholesterol. BMJ. 1994;308:373-9.
40. **Law MR, Wald NJ, Thompson SG.** By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischemic heart disease? BMJ. 1994;308:367-72.
41. **Rossouw JE, Lewis B, Rifkind BM.** The value of lowering cholesterol after myocardial infarction. N Engl J Med. 1990;323:1112-9.
42. Clofibrate and niacin in coronary heart disease. JAMA. 1975;231:360-81.
43. **Muldoon MF, Manuck SB, Matthews KA.** Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. BMJ. 1990;301:309-14.
44. **Holme I.** Relation of coronary heart disease incidence and total mortality to plasma cholesterol reduction in randomised trials: use of meta-analysis. Br Heart J. 1993;69(1 Suppl):S42-7.
45. **Law MR, Wald NJ, Thompson SG.** By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? BMJ. 1994;308:367-72.
46. **Law MR, Thompson SG, Wald NJ.** Assessing possible hazards of reducing serum cholesterol. BMJ. 1994;308:373-9.
47. **Gordon DJ.** Cholesterol lowering and total mortality. In: Rifkind BM, ed. Contemporary Issues in Cholesterol Lowering: Clinical and Population Aspects. New York: Marcel Dekker; 1994.
48. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet. 1994;344: 1383-9.
49. **Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, et al.** Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomized drug trials in their epidemiological context. Lancet. 1990;335:827-38.
50. Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. Hypertension Detection and Follow-up Program Cooperative Group. JAMA. 1979;242:2562-71.
51. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). SHEP Cooperative Research Group. JAMA. 1991;265: 3255-64.
52. **Furberg CD, Berglund G, Manolio TA, Psaty BM.** Overtreatment and undertreatment of hypertension. J Intern Med. 1994;235:387-97.
53. **Psaty BM, Heckbert SR, Koepsell TD, Siscovick DS, Lemaitre R, Smith NL, et al.** The risk of incident myocardial infarction associated with antihypertensive drug therapies [Abstract]. Circulation. 1996;91:925.
54. **Held PH, Yusuf S, Furberg CD.** Calcium channel blockers in acute myocardial infarction and unstable angina: an overview. BMJ. 1989;299:1187-92.
55. **Moertel CG.** Improving the efficiency of clinical trials: a medical perspective. Stat Med. 1984;3:455-68.
56. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. J Clin Oncol. 1992;10:896-903.
57. **Sande MA, Carpenter CC, Cobbs CG, Holmes KK, Sanford JD.** Antiretroviral therapy for adult HIV-infected patients. Recommendations from a state-of-the-art conference. National Institute of Allergy and Infectious Diseases State-of-the-Art Panel on Anti-Retroviral Therapy for Adult HIV-Infected Patients. JAMA. 1993;270:2583-9.
58. **Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, Muhs JM, et al.** Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med. 1990;322:802-9.
59. **Riggs BL, Seeman E, Hodgson SF, Taves DR, O'Fallon WM.** Effect of the fluoride/calcium regimen on vertebral fracture occurrence in postmenopausal osteoporosis. Comparison with conventional therapy. N Engl J Med. 1982; 306:446-50.

60. A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. The International Chronic Granulomatous Disease Cooperative Study Group. N Engl J Med. 1991;324:509-16.
61. **Comstock GW.** Identification of an effective vaccine against tuberculosis. Am Rev Respir Dis. 1988;138:479-80.
62. Urokinase-Streptokinase embolism trial. Phase 2 results. A cooperative study. JAMA. 1974;229:1606-13.
63. The Thrombolysis in Myocardial Infarction (TIMI) trial. Phase I findings. TIMI Study Group. N Engl J Med. 1985;312:932-6.
64. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio Della Streptochinasi nell'Infarto Miocardico (GISSI). Lancet. 1986;1:397-402.
65. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet. 1988;2: 349-60.
66. GISSI-2: a factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12,490 patients with acute myocardial infarction. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. Lancet. 1990;336:65-71.
67. In-hospital mortality and clinical course of 20,891 patients with suspected acute myocardial infarction randomised between alteplase and streptokinase with or without heparin. The International Study Group. Lancet. 1990;336: 71-5.
68. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO Investigators. N Engl J Med. 1993; 329:673-82.
69. **Fontana RS, Sanderson DR, Taylor WF, Woolner LB, Miller WE, Muhm JR, et al.** Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic Study. Am Rev Respir Dis. 1984;130:561-5.
70. **Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE, Muhm JR.** Lung cancer screening: the Mayo program. J Occup Med. 1986;28:746-50.
71. **Berson EL, Rosner B, Sandberg MA, Hayes KC, Nicholson BW, Weigel-DiFranco C, et al.** A randomized trial of vitamin A and vitamin E supplementation for retinitis pigmentosa. Arch Ophthalmol. 1993;111:761-72.
72. **Massof RW, Finkelstein D.** Supplemental vitamin A retards loss of ERG amplitude in retinitis pigmentosa [Editorial]. Arch Ophthalmol. 1993;111: 751-4.
73. Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease: a clinical trial. Nocturnal Oxygen Therapy Trial Group. Ann Intern Med. 1980;93:391-8.
74. **Hujoel PP, DeRouen TA.** A survey of endpoint characteristics in periodontal clinical trials published 1988-1992, and implications for future studies. J Clin Periodontol. 1995;22:397-407.
75. **Hujoel PP, DeRouen TA, Leroux B, Powell LV, Kiyak HA.** The validity of probing attachment levels as a surrogate for tooth loss in the elderly [Abstract]. J Dent Res. 1996;75:370.
76. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Survival data. Circulation. 1983;68:939-50.
77. **Lin DY, Fleming TR, DeGruttola V.** Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med. 1996; [In press].
78. **Freedman LS, Graubard BL, Schatzkin A.** Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992;11:167-78.
79. **DeGruttola V, Fleming T, Lin DY, Coombs R.** Validating surrogate markers—are we being naive? J Infect Dis. [In press].

"I presume you're tired after the long ride."

Turning her head at this, she answered solemnly: "I'm a great deal sicker than you think."

Her words fell on his ear with a great shock of wonder. He had often heard her pronounce them before—what if at last they were true?

He advanced a step or two into the dim room. "I hope that's not so, Zeena," he said.

She continued to gaze at him through the twilight with a mein of wan authority, as of one consciously singled out for a great fate. "I've got complications," she said.

Ethan knew the word for one of exceptional import. Almost everybody in the neighborhood had "troubles," frankly localized and specified; but only the chosen had "complications." To have them was in itself a distinction, though it was also, in most cases, a death warrant. People struggled on for years with "troubles," but they almost always succumbed to "complications."

Edith Wharton
Ethan Frome

Submitted by:
Joel E. Gallant, MD, MPH
Johns Hopkins University
Baltimore, MD 21287-6220

Submissions from readers are welcomed. If the quotation is published, the sender's name will be acknowledged. Please include a complete citation, as done for any reference.—*The Editor*