Users' Guides to the Medical Literature

XIX. Applying Clinical Trial Results
A. How to Use an Article Measuring the Effect of an Intervention on Surrogate End Points

Heiner C. Bucher, MD, MPH
Gordon H. Guyatt, MD, MSc
Deborah J. Cook, MD, MSc
Anne Holbrook, MD, MSc
Finlay A. McAlister, MD
for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

You are a physician seeing a 62-year-old woman with postmenopausal osteoporosis. Her bone mineral density, as measured by dual-energy x-ray absorptiometry, is 2.5 SDs below the mean value in premenopausal women. Although she does not have back pain, a spinal radiograph shows an old vertebral fracture. The patient has not yet experienced problems as a result of her vertebral fracture, but she is disturbed by the prospect that she may end up like her mother whose osteoporotic fractures have resulted in severe, long-term back pain.

The patient has reflux esophagitis and a past endoscopy revealed nonspecific gastritis. A specialist had prescribed alendronate, which the patient had to stop taking after several weeks because of dyspepsia. She searched the Web and discovered a new drug, raloxifene, and wonders whether this drug might be an alternative. You know that this drug has been licensed for the prevention of postmenopausal osteoporosis. You promise to examine the literature and to get back to her.

See also pp 786 and 790.

THE SEARCH

Using MEDLINE you identify a study of raloxifene for the treatment of osteoporosis demonstrating an effect on bone mineral density.[1] You are wondering whether this warrants administration to lower your patient's risk of osteoporotic fracture.

INTRODUCTION

Ideally, clinicians making treatment decisions should refer to methodologically strong clinical trials examining the impact of therapy on clinically important outcomes. By clinically important outcomes we mean outcomes that are important to patients: health-related quality of life, morbid end points such as stroke or myocardial infarction, or death. Often, however, conducting these trials requires such a large sample size, or long-term patient follow-up, that researchers or drug companies look for alternatives. Substituting surrogate end points for the target event allows conduct of shorter and smaller trials, thus offering an apparent solution to the dilemma.

A surrogate end point may be defined as "a laboratory measurement or a physical sign used as a substitute for a clinically meaningful end point that measures directly how a patient feels, functions or survives."[2] Surrogate end points include physiologic variables (such as bone mineral density as a surrogate for long-bone fractures, blood pressure for stroke, low-density lipoprotein cholesterol levels for myocardial infarction, and CD4 cell count for acquired immunodeficiency syndrome [AIDS] and AIDS-related mortality) or measures of subclinical disease (such as degree of atherosclerosis on coronary angiography).

The use of surrogate end points is indispensable for drug evaluation in phase 2 and early phase 3 trials geared to establishing a drug's promise of benefit. In many countries, companies may obtain drug approval by demonstrating a positive impact on surrogate end points. The use of surrogate end points for regulatory purposes reflects drug approval decisions that regulators must make in the face of public health exigencies.

Reliance on surrogate end points may be beneficial or harmful. On the one hand, use of the surrogate end point

Author Affiliations: Medizinische Universitäts-Poliklinik, Kantonsspital Basel, Basel, Switzerland (Dr Bucher); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario (Drs Guyatt, Cook, and Holbrook); and Division of General Internal Medicine, University of Alberta Hospital, Edmonton (Dr McAlister).

The original list of members (with affiliations) appears in the first article of the series (*JAMA*. 1993; 270:2093-2095). A list of new members appears in the 10th article of the series (*JAMA*.1996;275:1435-1439). The following members of the Evidence-Based Medicine Working Group contributed to this article: Antonio Dans, MD, Leonilla Dans, MD, Pat Brill-Edwards, MD, Daren Heyland, MD, Les Irwig, MBBCh, PhD, FFPHM, Roman Jaeschke, MD, MSc, Hui Lee, MD, MSc, Mitchell Levine, MD, MSc, Virginia Moyer, MD, MPH, and David Naylor, MD, DPhil.
Corresponding Author and Reprints: Gordon H. Guyatt, MD, MSc, McMaster University Health Sciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5.
Users' Guides to the Medical Literature Section Editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

Table 1. Users' Guide for a Surrogate End Point Trial

Are the results valid?
• Necessary, but not sufficient: is there a strong, independent, consistent association between the surrogate end point and the clinical end point?
• Is there evidence from randomized trials in other drug classes that improvement in the surrogate end point has consistently led to improvement in the target outcome?*
• Is there evidence from randomized trials in the same drug class that improvement in the surrogate end point has consistently led to improvement in the target outcome?*
What were the results?
• How large, precise, and lasting was the treatment effect? Effect should be large, precise, and lasting to consider a surrogate trial as possible basis for offering patients the intervention.
Will the results help me in caring for my patients?
• Are the likely treatment benefits worth the potential harms and costs? Offer intervention on basis of surrogate data only if patient's risk of the target outcome is high, patient places a high value on avoiding the target outcome, and if there are no satisfactory alternative therapies.

*Answers to one or both of these questions should be "yes" for surrogate trial to be an adequate guide for clinical action.

may lead to the rapid and appropriate dissemination of new treatments. For example, the Food and Drug Administration's decision to approve new antiretroviral drugs based on information from trials using surrogate end points recognized the enormous need for effective therapies for patients with human immunodeficiency virus (HIV) infection. Subsequently, several of these drugs have proved effective in randomized trials focusing on clinically important outcomes.[3-6]

On the other hand, reliance on surrogate end points may lead to excess morbidity and mortality. For example, while cardiac inotropes may improve short-term cardiac hemodynamic function in patients with heart failure, randomized clinical trials have demonstrated excess mortality with a number of these agents.[7] In particular, flosequinan was widely prescribed after its release, but had to be withdrawn after a trial revealed its deleterious effects on survival.[8]

How are clinicians to distinguish between these 2 situations? Surrogate outcome will be consistently reliable only if there is a causal connection between change in surrogate and change in the clinically important outcome. Thus, the surrogate must be in the causal pathway of the disease process and an intervention's entire effect on the clinical outcome of interest should be fully captured by a change in the surrogate. This Users' Guide builds on previous discussions of how one can establish a causal relationship[9] and presents an approach to critical appraisal of studies using surrogate end points and application of their results to manage individual patients.

As our discussion will make evident, the clinician needs to assess far more than a single study to make the decision about the adequacy of a surrogate. Evaluation may require a comprehensive review of observational studies of the relationship between the surrogate and the target, and of some or all of the randomized trials that have evaluated treatment impact on both the surrogate and the target. While most clinicians would hesitate to conduct such an investigation, our guidelines will allow them to evaluate the arguments made by experts or the pharmaceutical industry for prescribing treatments on the basis of their effect on surrogate end points.

THE GUIDES

In this guide, we follow the framework of previous articles in the series[10] and ask 3 sorts of questions: are the results valid; what were the results; and will the results help me in caring for my patients? (TABLE 1). When we consider the validity of a surrogate, we must address 2 issues. First, to be consistently reliable, the surrogate must be in the causal pathway from the intervention to the outcome. Second, in considering a particular intervention, we must be confident that there are no important effects of that intervention on the outcome of interest that are not mediated through, or captured by, the surrogate. Our guides for validity (Table 1) bear directly on these 2 issues.

Are the Results Valid? Is There a Strong, Independent, Consistent Association Between the Surrogate End Point and the Clinical End Point?

To provide a valid substitute for an important target outcome, the surrogate must be associated or correlated with that target. In general, researchers choose surrogate end points because they have found a correlation between a surrogate and a target outcome in observational studies, and their understanding of the biology makes it plausible that changes in the surrogate will invariably lead to changes in the important outcome. The stronger the association, the more likely the causal link between the surrogate and the target. The strength of an association is reflected in statistics such as relative risk (RR) or odds ratio. We have presented a full discussion of statistics reflecting the strength of association in another article.[11] Many biologically plausible surrogates are only weakly associated with clinically important outcomes. For example, measures of respiratory function in patients with chronic lung disease, or conventional exercise tests in patients with heart and lung disease, are only weakly correlated with capacity to undertake activities of daily living.[12,13] When correlations are low, the surrogate is likely to be a poor substitute for the target outcome.

In addition to the strength of the association, one's confidence in the validity of the association depends on whether it is consistent across different studies and after adjustment for known confounders. For example, ecologic studies such as the Seven Countries Study[14] suggested a strong correlation between serum cholesterol levels and coronary heart disease mortality even after adjusting for other predictors such as age, smoking, and systolic blood pressure. Subsequent cohort studies confirmed this association and suggested that long-term reductions in serum cholesterol levels of 0.6 mmol/L (23 mg/dL) would lower the risk of coronary heart disease by approximately 30%. When a surrogate is associated with an outcome after ad-

justing for multiple other potential prognostic factors we call the association *independent*.

Similarly, cohort studies have consistently revealed that a single measurement of plasma viral load predicts the subsequent risk of AIDS or death in patients infected with HIV.[15-20] For example, in 1 study the proportion of patients that progressed to AIDS after 5 years in the lowest through the highest quartiles of viral load was 8%, 26%, 49%, and 62%, respectively.[20] Moreover, this association retained its predictive power after adjustment for other potential predictors such as CD4 cell count.[15-19]

Returning to the scenario, you are wondering if you can substitute bone mineral density for fractures or health-related quality of life in considering whether to recommend raloxifene. A large cohort study investigated risk factors for hip fracture.[21] Postmenopausal women with a calcaneal bone density in the highest third had a hip fracture rate of 9.4/1000 woman-years while women in the middle and lowest third had a fracture rate per 1000 woman-years of 14.7 and 27.3, respectively. Furthermore, after considering other risk factors for osteoporotic hip fractures including maternal history of hip fracture, previous fractures from any site, poor self-rated health, use of long-acting benzodiazepines, impaired visual function, and reduced physical activity, bone mineral density continued to predict the risk of hip fracture.[21] These findings are consistent across studies looking at the association between bone density and fracture risk.[22,23] Thus, bone mineral density is a moderately strong, independent predictor of fracture, and meets our first criterion for an acceptable surrogate end point.

While meeting this first criterion is necessary, it is not sufficient to support reliance on a surrogate outcome. As we will emphasize below (Table 1), before offering an intervention on the basis of effects on a surrogate outcome, the clinician should note a consistent relationship between surrogate and target in randomized trials; the effect of the intervention on the surrogate must be large, precise, and lasting, and the benefit-risk trade-off must be clear.

Is There Evidence From Randomized Trials in Other Drug Classes That Improvement in the Surrogate End Point Has Consistently Led to Improvement in the Target Outcome?

Given the possibility of effects unrelated to the surrogate end point, pathophysiologic studies, ecological studies, and cohort studies are insufficient to establish that the link between surrogate and clinically important outcomes is ironclad. We can confidently rely on surrogate end points only when long-term randomized trials have consistently demonstrated that modification of the surrogate is associated with concomitant modifications in the target outcome of interest. For example, although ventricular ectopic beats are associated with adverse prognosis in patients with myocardial infarction[24] and class 1 antiarrhythmic agents effectively suppress ventricular arrhythmias in animals and humans,[25] these drugs have proved to increase mortality when evaluated in randomized trials.[26] In this case, reliance on the surrogate end point of suppression of nonlethal arrhythmias led to the deaths of tens of thousands of patients.[27]

The treatment of heart failure provides another instructive example. Trials of angiotensin-converting enzyme inhibitors in heart failure treatment have demonstrated parallel increases in exercise capacity[28-31] and decreases in mortality,[32] suggesting that clinicians may be able to rely on exercise capacity as a valid surrogate. Milrinone[33] and epoprostenol[34] have both demonstrated improved exercise tolerance in patients with symptomatic heart failure. However, when these drugs were evaluated in randomized controlled trials both showed an increase in cardiovascular mortality that in one instance was statistically significant,[35] and in the second case led to the early termination of the study.[36] Thus, exercise tolerance is inconsistent in predicting improved mortality and is therefore an unsatisfactory substitute. Other suggested surrogate end points in heart failure have included ejection fraction, heart-rate variability, and markers of autonomic function.[37] The dopaminergic agent ibopamine positively influences all 3 surrogate end points, and yet a randomized trial demonstrated that the drug increases mortality in heart failure.[38]

An example of a surrogate end point is CD4 cell count, which has been validated in randomized trials. A number of trials comparing different classes of antiretroviral therapies have demonstrated that patients randomized to more potent drug regimens had higher CD4 cell counts and were less likely to progress to AIDS or death.[6,39] While there is no guarantee that the next trial using a different class of drugs will show the same pattern, these results greatly strengthen our inference that if therapy for HIV infection increases the CD4 count, a reduction in AIDS-related mortality will result.

Returning to our scenario, trials of etidronate[40,41] and alendronate[42] for the prevention of osteoporotic fractures in postmenopausal women have shown parallel increases in bone mineral density and reduced incidences of new vertebral fractures. This would suggest that clinicians might rely on bone density to evaluate new drugs in osteoporosis in making the assumption that if they saw increases in bone density, decreases in fractures would follow.

However, another secondary prevention trial in postmenopausal women using sodium fluoride showed divergent results.[43] Although sodium fluoride increased bone mineral density at the lumbar spine by 35% over 5 years, more vertebral and nonvertebral fractures occurred in the intervention group than in the placebo group (163 and 72 in 101 women with sodium fluoride vs 136 and 24 in 101 women with placebo). In another randomized trial, fluoride again showed a large increase in bone density without any change in fracture rate.[44] Inferences on the basis of unchanged bone density may also be problematic. A study of calcium and vitamin D in the elderly showed virtually no change in bone density, but a reduction in fracture risk of approximately 50%.[45] Thus, increase in

bone mineral density as a surrogate end point has shown an inconsistent relationship to osteoporotic fractures.

Is There Evidence From Randomized Trials in the Same Drug Class That Improvement in the Surrogate End Point Has Consistently Led to Improvement in the Target Outcome?

Clinicians are in a stronger position to rely on surrogate end points if the new drug they are considering is from a class of drugs in which the relationship between changes in the surrogate and changes in the target has been verified in randomized trials. For instance, thiazide diuretics and β-blockers have both been shown to reduce blood pressure and clinically important outcomes such as stroke in patients with hypertension. Thus, we would be much more comfortable relying on reduction in blood pressure to justify administering a new β-blocker or thiazide diuretic than to justify offering a novel antihypertensive agent from another class.[46]

For example, although 1 dihydropyridine calcium channel blocker has been shown to reduce clinically important outcomes in patients with hypertension,[47] 4 other trials have shown that these agents are less efficacious than thiazides or angiotensin-converting enzyme inhibitors in preventing hard clinical end points despite exerting similar degrees of blood pressure lowering.[48-51]

We will consider the example of cholesterol reduction as a surrogate for cardiovascular outcomes such as myocardial infarction and death in part B of this Users' Guide.[52] Briefly, several large trials of primary and secondary prevention of coronary heart disease with statins have consistently shown that these drugs reduce cardiovascular outcomes.[53]

We could therefore make the assumption that a new statin with a similar low-density lipoprotein cholesterol–lowering potency may also reduce clinically important outcomes. However, we would be reluctant to generalize to another class of lipid-lowering agents since trials of 1 such class (the fibrates) have shown that these drugs reduce the incidence of myocardial infarction but increase the risk of mortality from other causes (with no impact on overall mortality).[53-55]

These examples highlight the point we made earlier: confidence in a surrogate outcome depends on the assumption that the treatment captures any relationship between the treatment and the outcome.[56,57] This assumption can be violated in 2 ways. First, treatment may have a beneficial mechanism of effect on the outcome independent of its effect on the surrogate. For instance, 1 explanation for the superior effect of angiotensin-converting enzyme inhibitors vs calcium antagonists on clinically important outcomes is that angiotensin-converting enzyme inhibition has biological effects independent of lowering blood pressure that reduce risk of stroke or death and that calcium antagonists do not share these effects.

Second, treatment may have deleterious effects on the outcome that are not mediated through the surrogate. Mortality-increasing effects of fibrates rather than inability to lower morbidity and mortality through cholesterol reduction probably explain the lack of effect of fibrates on clinically important outcomes. That such additional effects are less likely across classes of drugs than within classes is what makes us more inclined to rely on within-class evidence from surrogate outcomes.

This criterion is complicated by the variable definitions of drug class. A manufacturer of a drug related to a class of agents with a consistently positive association between modification of a surrogate end point and modification of the target (such as a β-blocker) will naturally argue for a broad definition of class. Manufacturers of agents that are related to drugs with known or suspected adverse effects on target events (clofibrate, or some calcium antagonists) are likely to argue, on the other hand, that the chemical or physiological connection is not sufficiently close to consider the new drug to be in the same class as the harmful agent. Part B will address these issues more fully.[52]

Returning to the scenario, we have established that because of the inconsistent relationship between increase in bone mineral density and fracture reduction we would be reluctant to offer patients a new antiosteoporotic agent solely on the basis of evidence of its effect on the surrogate end point. Raloxifene, the drug we are considering for our patient, is a nonsteroidal benzothiophene, a selective estrogen-receptor modulator representing a new class of drugs for the prevention of osteoporosis-related bone fractures. Thus, it is likely that the mechanisms of action will be considerably different from bisphosphonates and the conclusion that similar reductions in loss of bone density will lead to parallel reductions in clinical fractures is questionable. In TABLE 2, we apply our validity criteria to a number of controversial examples of the use of surrogate end points.

What Were the Results? How Large, Precise, and Lasting Was the Treatment Effect?

We are interested not only in whether an intervention alters a surrogate end point, but also in the magnitude, precision, and duration of the effect. If an intervention shows large reductions in the surrogate end point, the 95% confidence intervals (CIs) around those large reductions are narrow, and the effect persists over a sufficiently long period, our confidence that the target outcome will be favorably affected increases. Positive effects that are smaller, with wider CIs, and shorter duration of follow-up leave us less confident.

We have already cited evidence suggesting that CD4 cell counts may be an acceptable surrogate for mortality in patients with HIV infection. A randomized controlled trial of immediate vs delayed zidovudine therapy in HIV-infected asymptomatic individuals declared a positive result for immediate therapy, largely on the basis of a greater proportion of treated patients with CD4 cell counts above 435×10^6/L at a median follow-up of 1.7 years.[58] Subsequently, the Concorde study addressed the same question in a randomized trial

with a median follow-up of 3.3 years.[59] The Concorde investigators found a continuous decline in CD4 cell counts in both treated and control groups, but the median difference of 30 × 10^6/L in favor of treated patients at study termination was statistically significant. However, the study showed no effect of zidovudine in terms of reduced progression to AIDS or death. The median CD4 cell count difference was insufficient to have an impact on clinically important outcomes. The Concorde authors made the following conclusion: the small, but highly significant persistent difference in CD4 cell counts between the groups was not translated into a significant clinical benefit and "called into question the uncritical use of CD4 cell counts as a surrogate endpoint." Had the Concorde analysis showed significantly shorter times to reach a CD4 cell count of 350 × 10^6/L in the control group and been regarded as fundamental, the trial might have been stopped early with a false-positive result.

Returning to our scenario, the trial of raloxifene in women with osteoporosis demonstrated that after 2 years of treatment, raloxifene-treated patients in the group receiving the highest dosage showed an increase in bone mineral density at the lumbar spine of 2.2% (SE, 0.3%) compared with a slight decrease in the control group 0.8% (SE, 0.3%). This difference in change over time was statistically significant ($P<.03$). Ideally, the investigators would have provided us with a CI around the 3% difference in percentage change in bone mineral density in the treatment and control groups. As we will illustrate when we consider weighing benefits and harms, the magnitude of the effect on the surrogate may (or may not) help us estimate the size of a possible affect on the target outcome.

Will the Results Help in Caring for My Patients?

The questions clinicians should ask themselves in applying the results are the same ones we have suggested for any issue of therapy or prevention [60] and elaborated on in our Users' Guide regarding applicability.[61] These 3 questions have to do with whether the results can be applied to your patient's care, whether all important outcomes were considered, and whether the likely benefits are worth the down sides of treatment.

"Can the results be applied to my patient's care" refers to the extent to which the patient before you is similar to those who participated in the published studies under consideration, and the extent to which the therapy, and the associated technologies for monitoring and responding to complications, are available in your setting. "Were all important outcomes considered" relates to the focus of this Users' Guide, and all the issues we have raised thus far: was the primary outcome really the one in which patients will be interested?

This second criterion also draws issues of adverse intervention effects to our attention. Applying the third cri-

Table 2. Selected Examples of Applied Validity Criteria for the Critical Evaluation of Studies Using Surrogate End Points

	Criterion				
Types of Intervention	Is There a Strong, Independent, Consistent Association Between the Surrogate End Point and the Clinical End Point?	Is There Evidence From Randomized Trials in Other Drug Classes That Improvement in the Surrogate End Point Has Consistently Led to Improvement in the Target Outcome?	Is There Evidence From Randomized Trials in the Same Drug Class That Improvement in the Surrogate End Point Has Consistently Led to Improvement in the Target Outcome?	Surrogate End Point	End Point
Nonsteroidal benzothiophene Raloxifene[1]	Yes[21-23]	No[43,44]	No[1,62]	Bone mineral density	Osteoporotic fractures
Protease inhibitor* Nelfinavir[63]	Yes[15-19]	Yes[64-66]	Yes[6,39]	Human immunodeficiency virus plasma load	Acquired immunodeficiency syndrome or death
Reverse transcriptase inhibitor Abacavir[67]	Yes[15-19]	Yes[65,66,68]	Yes[64,68]	Human immunodeficiency virus viral plasma load	Acquired immunodeficiency syndrome or death
Protease inhibitor* Nelfinavir[63]	Yes[15-19]	Yes[64-66]	Yes[6,39]	CD4 cell count	Acquired immunodeficiency syndrome or death
Reverse transcriptase inhibitor Abacavir[67]	Yes[15-19]	Yes[65,66,68]	Yes[64,68]	CD4 cell count	Acquired immunodeficiency syndrome or death
Antihypertensive drugs Dihydropyridine calcium antagonist New thiazide diuretic	Yes[69,70] Yes[69,70]	Yes[71] Yes[71]	No[48-51] Yes[71]	Blood pressure reduction	Stroke, myocardial infarction, cardiovascular mortality
Antilipidemic drugs Atorvastatin[72,73] Bezafibrate[75,76]	Yes[14,74] Yes[14,74]	No[53] No[53]	Yes[53] No[53]	Cholesterol or low-density lipoprotein cholesterol	Myocardial infarction, death from myocardial infarction

*In combination therapy with 2 reverse transcriptase inhibitors.

terion, judging whether the benefits are worth the down sides of treatment, presents particular challenges when investigators have focused on surrogate end points, and we will discuss this criterion in some detail.

Are the Likely Treatment Benefits Worth the Potential Harms and Costs?

To know whether to offer a treatment to their patients, clinicians must be able to estimate the magnitude of the likely benefit. When the data available are limited to the effect on a surrogate end point, estimating the extent to which treatment will reduce clinically important outcomes becomes a challenge.

One approach is to extrapolate from 1 or more randomized trials assessing a related intervention in a similar patient population that provides both surrogate end point and clinical outcome data. For example, until recently there were little long-term data on the efficacy of lovastatin in reducing clinically important outcomes. However, one could extrapolate from short-term dose efficacy studies assessing the surrogate end point of cholesterol lowering. Thus, since 40 mg of lovastatin produced a similar degree of lowering of low-density lipoprotein cholesterol as 40 mg of pravastatin (31% vs 34% reduction) in the CURVES Study,[77] one could theorize that lovastatin would have similar long-term benefits to pravastatin. Subsequently, the AFCAPS/TexCAPS Trial (a 5-year trial assessing the efficacy of lovastatin in the primary prevention of ischemic heart disease)[78] confirmed that this agent had a beneficial profile similar to pravastatin (as determined by the 5-year, primary prevention WOSCOPS Trial)[79]: the RR reductions (and 95% CIs) for myocardial infarction were 40% (17%-57%) and 31% (17%-43%), respectively. However, this approach is likely to be seriously flawed when one is extrapolating from trials of another class of drugs.

Returning to our scenario, to estimate the magnitude of the fracture reduction we might expect with raloxifene (in which we have only surrogate end point data), we could (recognizing the limitations of this approach pointed out above) examine the results of randomized controlled trials of alendronate (a drug from a different class for which we have data on the same surrogate end point as well as clinical end points such as fracture reduction). While alendronate appears to improve vertebral bone density by 7.5% over 2 years (vs control),[42] raloxifene is associated with only a 3.0% improvement over the same time frame. A systematic overview of the alendronate trials[80] reported a 29% reduction in RR of nonvertebral fracture over 2 years. Only 1 trial looked at symptomatic vertebral fractures in women with decreased bone density and an existing vertebral fracture.[81] This study demonstrated an RR reduction of 55% with alendronate and suggested that our patient's risk over 3 years of a nonvertebral fracture would be approximately 15%; symptomatic vertebral fracture would be about 5%. Given the RR reductions with alendronate, one would need to treat approximately 25 women to prevent a nonvertebral fracture and 40 women to prevent a symptomatic vertebral fracture over a 3-year period.

Since the improvement in bone mineral density with raloxifene is at best 50% of the effect of alendronate, we would anticipate a considerably lower reduction in fracture risk with raloxifene. However, interim analysis of an ongoing raloxifene trial[62] reported a 46% RR reduction with this therapy (despite less of an increase in bone mineral density than seen with the alendronate trials). This serves to emphasize the dangers of extrapolating results across classes when it is uncertain that the effects on clinically important outcomes are mediated in the same fashion by the 2 comparison drugs.

In deciding whether the likely magnitude of the treatment effect warrants offering patients the intervention, clinicians must consider not only the uncertainty associated with that estimate, but the trade-off with potential toxic effects and costs of therapy. In addition, clinicians must ponder the consequences of not treating, and the available management alternatives. The deadly and usually relentless progression of HIV infection, and the paucity of alternative therapies, has contributed to the readiness of patients, clinicians, and regulatory agencies to accept evidence from surrogate end points in instituting novel therapies in patients infected with HIV. In osteoporosis, in which the consequences of the condition are less immediately devastating, and a variety of agents are available, the case for relying on surrogate end points is far less compelling.

RESOLUTION OF THE SCENARIO

We have found a strong, consistent, independent, and biologically plausible association between bone mineral density and vertebral and nonvertebral fractures. Randomized trials, however, have failed to show a consistent association between increased bone density and reduction in fracture across all drug classes.

Because our patient is at substantial risk of fracture over the short term, the number needed to treat to prevent both nonvertebral and vertebral fractures is moderate, as is the absolute benefit she might expect. Moreover, she is interested in longer-term fracture prevention, and her risk will grow over time. One might offer her alternative interventions, including hormone replacement therapy, calcium and vitamin D, bisphosphonates, or calcitonin.

While there is strong evidence from randomized trials supporting the use of bisphosphonates to decrease osteoporotic fractures, randomized trial data showing fracture reduction in populations similar to our patient with the other agents is limited. Our patient is concerned about her long-term risk. Raloxifene was well tolerated during this 2-year trial but no information is available about long-term adverse effects including cardiovascular disease, venous thromboembolism, breast and endometrial cancer, and menopausal symptoms. While a number of options (including a trial of etidronate, offering hormone replacement therapy, calcium and vitamin D, calcitonin, or suggesting only a bal-

anced diet and exercise) might be reasonable, ideally the clinician would subject these options to the same scrutiny applied to raloxifene.

Data indicating a reduction in fracture rate would greatly strengthen the case for including raloxifene as the preferred option. Just as you are about to see the patient (and, for us, just before this article went to press) you pick up a few of your latest editions of *JAMA* from the pile in the corner of your office, and find 2 highly relevant randomized trials.[82,83] The results show that, in women like your patient with a prevalent vertebral fracture, raloxifene decreased radiological vertebral fracture risk (for 60 mg: number needed to treat = 16 [RR, 0.7; 95% CI, 0.6-0.9]; and for 120 mg: number needed to treat = 10 [RR, 0.5; 95% CI, 0.4-0.7]), but did not decrease the incidence of nonvertebral fracture. In helping your patient to decide on the right course of action, you realize you will have to consider other effects of raloxifene: the *JAMA* articles also show a 76% RR reduction of breast cancer as detected by mammography (number needed to treat, 126), a 3-fold increase in the risk of venous thromboembolism, and an increased incidence of hot flashes, leg cramps, influenzalike syndromes, and peripheral edema.

When we use surrogate end points to make inferences about expected benefit, we are making assumptions regarding the link between the surrogate end point and the target outcome. We have outlined criteria clinicians can use to decide when these assumptions might be appropriate. Even if a surrogate end point meets all of these criteria, inferences about a treatment benefit may still prove misleading. Thus, treatment recommendations based on surrogate outcome effects can never be strong. Furthermore, difficulties in estimating the magnitude of effects on clinically important end points compromises economic analysis examining the cost-effectiveness of alternative management strategies.

These considerations emphasize that waiting for randomized trials investigating the effect of the intervention on outcomes of unequivocal importance to patients is the only ironclad solution to the surrogate outcome dilemma. When clinicians must choose between alternative interventions, trials should make head-to-head comparisons between competing treatments rather than restricting comparisons of treatment to control or placebo. We expand on this issue in Part B of this Users' Guide. However, when patients' risk of serious morbidity or mortality are high, this "wait-and-see" strategy may pose problems for many patients and their physicians.

We encourage clinicians to critically question therapeutic interventions in which the only proof of efficacy is from surrogate end point data. When the surrogate end point meets all our validity criteria, the effect of the intervention on the surrogate end point is large, the patient's risk of the target outcome is high, the patient places a high value on avoiding the target outcome, and there are no satisfactory alternative therapies, clinicians can recommend therapy on the basis of randomized trials evaluating only surrogate end points. In other situations, clinicians must carefully consider the known adverse effects and cost of therapy, and the possibility of unanticipated adverse effects, before recommending an intervention solely on the basis of surrogate end point data.

Acknowledgment: We are grateful to Cliff Rosen, MD, for his helpful comments concerning the scenario and the associated discussion. Deborah Maddock provided invaluable coordination for the EBM Working Group in the development of this article.

REFERENCES

1. Delmas PD, Bjarnason NH, Mitlak BH, et al. Effects of raloxifene on bone mineral density, serum cholesterol concentrations, and uterine endometrium in postmenopausal women. *N Engl J Med.* 1997;337:1641-1647.
2. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, eds. *Clinical Measurement in Drug Evaluation.* New York, NY: John Wiley & Sons Inc; 1995:57.
3. Hammer SM, Katzenstein DA, Hughes MD, et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *N Engl J Med.* 1996;335:1081-1090.
4. Delta Coordinating Committee. Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. *Lancet.* 1996;348:283-291.
5. Saravolatz LD, Winslow DL, Collins G, et al. Zidovudine alone or in combination with didanosine or zalcitabine in HIV-infected patients with the acquired immunodeficiency syndrome or fewer than 200 CD4 cells per cubic millimeter. *N Engl J Med.* 1996;335:1099-1106.
6. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *N Engl J Med.* 1997;337:725-733.
7. Niebauer J, Coats AJ. Treating chronic heart failure: time to take stock. *Lancet.* 1997;349:966-967.
8. Massie BM, Berk MR, Brozena SC, et al. Can further benefit be achieved by adding flosequinan to patients with congestive heart failure who remain symptomatic on diuretic, digoxin, and an angiotensin converting enzyme inhibitor? results of the flosequinan-ACE inhibitor trial (FACET). *Circulation.* 1993;88:492-501.
9. How to read clinical journals, IV: to determine etiology or causation. *CMAJ.* 1981;124:985-990.
10. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature, I: how to get started. *JAMA.* 1993;270:2093-2095.
11. Guyatt G, Walter S, Shannon H, Cook D, Jaeschke R, Heddle N. Basic statistics for clinicians, IV: correlation and regression. *CMAJ.* 1995;152:497-504.
12. Guyatt GH, Thompson PJ, Berman LB, et al. How should we measure function in patients with chronic heart and lung disease? *J Chronic Dis.* 1985;38:517-524.
13. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest.* 1984;85:751-758.
14. Verschuren WM, Jacobs DR, Bloemberg BP, et al. Serum total cholesterol and long-term coronary heart disease mortality in different cultures. *JAMA.* 1995;274:131-136.
15. Mellors JW, Rinaldo CR Jr, Gupta P, et al. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science.* 1996;272:1167-1170.
16. Mellors JW, Kingsley LA, Rinaldo CR, et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann Intern Med.* 1995;122:573-579.
17. Ruiz L, Romeu J, Clotet B, et al. Quantitative HIV-1 RNA as a marker of clinical stability and survival in a cohort of 302 patients with a mean CD4 cell count of 300 × 10(6)/1. *AIDS.* 1996;10:F39-F44.
18. O'Brien TR, Blattner WA, Waters D, et al. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. *JAMA.* 1996;276:105-110.
19. Yerly S, Perneger TV, Hirschel B, et al. A critical assessment of the prognostic value of HIV-1 RNA levels and CD4+ cell counts in HIV-infected patients. *Arch Intern Med.* 1998;158:247-252.
20. Ho DD. Viral counts in HIV infection. *Science.* 1996;272:1124-1125.
21. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. *N Engl J Med.* 1995;332:767-773.
22. Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ.* 1996;312:1254-1259.
23. Huang C, Ross PD, Wasnich RD. Short-term and long-term fracture prediction by bone mass measurements. *J Bone Miner Res.* 1998;13:107-113.
24. Bigger JT Jr, Fleiss JL, Kleiger R, et al. The relationships among ventricular arrhythmias, left ventricular dysfunction, and mortality in the 2 years after myocardial infarction. *Circulation.* 1984;69:250-258.
25. McAlister FA, Teo KK. Antiarrhythmic therapies for the prevention of sudden cardiac death. *Drugs.* 1997;54:235-252.
26. Echt DS, Liebson PR, Mitchell LB, et al, and the Cardiac Arrhythmia Suppression Trial. Mortality and

morbidity in patients receiving encainide, flecainide, or placebo. *N Engl J Med.* 1991;324:781-788.
27. Moore TJ. *Deadly Medicine.* New York, NY: Simon & Schuster; 1995.
28. Drexler H, Banhardt U, Meinertz T, et al. Contrasting peripheral short-term and long-term effects of converting enzyme inhibition in patients with congestive heart failure: a double-blind, placebo-controlled trial. *Circulation.* 1989;79:491-502.
29. Lewis GR. Comparison of lisinopril versus placebo for congestive heart failure. *Am J Cardiol.* 1989; 63:12D-16D.
30. Giles TD, Fisher MB, Rush JE. Lisinopril and captopril in the treatment of heart failure in older patients. *Am J Med.* 1988;85:44-47.
31. Riegger GA. Effects of quinapril on exercise tolerance in patients with mild to moderate heart failure. *Eur Heart J.* 1991;12:705-711.
32. Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. *JAMA.* 1995;273:1450-1456.
33. Di Bianco R, Shabetai R, Kostuk W, et al. A comparison of oral milrinone, digoxin, and their combination in the treatment of patients with chronic heart failure. *N Engl J Med.* 1989;320:677-683.
34. Sueta CA, Gheorghiade M, Adams KF, et al, and the Epoprostenol Multicenter Research Group. Safety and efficacy of epoprostenol in patients with severe congestive heart failure. *Am J Cardiol.* 1995;75:34A-43A.
35. Packer M, Carver JR, Rodeheffer RJ, et al, for the Promise Study Research Group. Effect of oral milrinone on mortality in severe chronic heart failure. *N Engl J Med.* 1991;325:1468-1475.
36. Califf RM, Adams KF, McKenna WJ, et al. A randomized controlled trial of epoprostenol therapy for severe congestive heart failure. *Am Heart J.* 1997; 134:44-54.
37. Yee KM, Struthers AD. Can drug effects on mortality in heart failure be predicted by any surrogate measure? *Eur Heart J.* 1997;18:1860-1864.
38. Hampton JR, van Veldhuisen DJ, Kleber FX, et al. Randomised study of effect of ibopamine on survival in patients with advanced severe heart failure. *Lancet.* 1997;349:971-977.
39. Cameron DW, Heath-Chiozzi M, Danner S, et al, and the Advanced HIV Disease Ritonavir Study Group. Randomised placebo-controlled trial of ritonavir in advanced HIV-1 disease. *Lancet.* 1997;351:543-549.
40. Watts NB, Harris ST, Genant HK, et al. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *N Engl J Med.* 1990;323:73-79.
41. Storm T, Thamsborg G, Steiniche T, Genant HK, Sorensen OH. Effect of intermittent cyclical etidronate therapy on bone mass and fracture rate in women with postmenopausal osteoporosis. *N Engl J Med.* 1990;322:1265-1271.
42. Liberman UA, Weiss SR, Broll J, et al, and the Alendronate Phase III Osteoporosis Treatment Study Group. Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med.* 1995;333:1437-1443.
43. Riggs BL, Hodgson SF, O'Fallon WM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *N Engl J Med.* 1990;322:802-809.
44. Meunier PJ, Sebert J-L, Reginster J-Y, et al. Fluoride salts are no better at preventing new vertebral fractures than calcium-vitamin D in postmenopausal osteoporosis. *Osteoporos Int.* 1998;8:4-12.
45. Dawson-Hughes B, Harris SS, Krall EA, Dallal GE. Effect of calcium and vitamin D supplementation on bone density in men and women 65 years of age and older. *N Engl J Med.* 1977;337:670-676.
46. McAlister FA, Straus S, Sackett DL. Randomized clinical trials of antihypertensive drugs: all that glitters is not gold. *CMAJ.* 1998;159:488-490.
47. Staessen JA, Fagard R, Thijs L, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. *Lancet.* 1997;350:757-764.
48. Psaty BM, Siscovick DS, Weiss NS, et al. Hypertension and outcomes research: from clinical trials to clinical epidemiology. *Am J Hypertens.* 1996;9:178-183.
49. Borhani NO, Mercuri M, Borhani PA, et al. Final outcome results of the multicenter isradipine diuretic atherosclerosis study (MIDAS): a randomized controlled trial. *JAMA.* 1996;276:785-791.
50. Tatti P, Pahor M, Byington RP, et al. Outcome results of the fosinopril versus amlodipine cardiovascular events randomized trial (FACET) in patients with hypertension and NIDDM. *Diabetes Care.* 1998;21:597-603.
51. Estacio RO, Jeffers BW, Hiatt WR, Biggerstaff SL, Gifford N, Schrier RW. The effect of nisoldipine as compared with enalapril on cardiovascular outcomes in patients with non-insulin-dependent diabetes and hypertension. *N Engl J Med.* 1998;338:645-652.
52. McAlister FA, Laupacis A, Wells GA, Sackett DL, for the Evidence-Based Medicine Working Group. Users' guide to the medical literature, XIX: applying clinical trial results part B: guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA.* In press.
53. Bucher HC, Griffith LE, Guyatt GH. Systematic review on risk and benefit of different cholesterol lowering interventions. *Arterioscler Thromb Vasc Biol.* 1999;19:187-195.
54. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentration and mortality. *BMJ.* 1990;301:309-314.
55. Smith GD, Song F, Sheldon TA. Cholesterol lowering and mortality. *BMJ.* 1993;306:1367-1373.
56. Prentice RL. Surrogate endpoints in clinical trials. *Stat Med.* 1989;8:431-440.
57. Fleming TR. Surrogate markers in AIDS and cancer trials. *Stat Med.* 1994;13:1423-1435.
58. Cooper DA, Gatell JM, Kroon S, et al, and the European-Australian Collaborative Group. Zidovudine in persons with asymptomatic HIV infection and CD4+ cell counts greater than 400 per cubic millimeter. *N Engl J Med.* 1993;329:297-303.
59. Concorde Coordinating Committee. Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *Lancet.* 1994;343:871-881.
60. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention A: are the results of the study valid? *JAMA.* 1993;270:2598-2601.
61. Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, XIV: how to decide on the applicability of clinical trial results to your patient. *JAMA.* 1998;279:545-549.
62. Ettinger B, Black D, Cummings S, et al. Raloxifene reduces the risk of incident vertebral fractures: 24 month interim analyses. *Osteoporos Int.* 1998;8 (suppl 3):11.
63. Clendeneninn N, Quart B, Anderson R, Knowles M, Chang Y. Analysis of long-term virologic data from the viracept (nelfinavir) 511 protocol using 3 HIV-RNA assays. In: Abstracts from the 5th Conference on Retroviruses and Opportunistic Infections; Chicago, Ill; 1998.
64. Brun-Vezinet F, Boucher C, Loveday C, et al, and the Delta Virology Working Group and Coordinating Committee. HIV-1 viral load, phenotype, and resistance in a subset of drug-naive participants from the Delta trial: the national virology groups. *Lancet.* 1997; 350:983-990.
65. Montaner JS, Reiss P, Cooper D, et al. A randomized, double-blind trial comparing combinations of nevirapine, didanosine, and zidovudine for HIV-infected patients: the Netherlands, Canada, and Australia (INCAS) study. *JAMA.* 1998;279:930-937.
66. CEASAR Coordinating Committee. Randomised trial of addition of lamivudine or lamivudine plus loviride to zidovudine-containing regimens for patients with HIV-1 infection: the CAESAR trial. *Lancet.* 1997;349:1413-1421.
67. Fischl M, Greenberg S, Clumeck N, et al. Safety and activity of abacavir (1592, ABC) with 3TC/ZDV in antiretroviral naive subjects. In: Abstracts from the 12th World AIDS Conference: Geneva, Switzerland; June 28-July 3, 1998.
68. Katzenstein DA, Hammer SM, Hughes MD, et al, and the AIDS Clinical Trials Group Study 175 Virology Study Team. The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. *N Engl J Med.* 1996;335:1091-1098.
69. Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease, part 2: short-term reductions in blood pressure; overview of randomised drug trials in their epidemiological context. *Lancet.* 1990;335:827-838.
70. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease, part 1: prolonged differences in blood pressure; prospective observational studies corrected for the regression dilution bias. *Lancet.* 1990;335:765-774.
71. Psaty BM, Smith NL, Siscovick DS, et al. Health outcomes associated with antihypertensive therapies used as first-line agents: a systematic review and meta-analysis. *JAMA.* 1997;277:739-745.
72. Heinonen TM, Stein E, Weiss SR, et al. The lipid-lowering effects of atorvastatin, a new HMG-coA reductase inhibitor: results of a randomized, double-masked study. *Clin Ther.* 1996;18:853-863.
73. Bakker-Arkema RG, Davidson MH, Goldstein RJ, et al. Efficacy and safety of a new HMG-coA reductase inhibitor, atorvastatin, in patients with hypertriglyceridemia. *JAMA.* 1996;275:128-133.
74. 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;308:367-372.
75. Winocour PH, Durrington PN, Bhatagnar D, et al. The effect of bezafibrate on very low density lipoprotein (VLDL), intermediate density lipoprotein (IDL), and low density lipoprotein (LDL) composition in type 1 diabetes associated with hypercholesterolaemia or combined hyperlipidaemia. *Atherosclerosis.* 1992;93:83-94.
76. Jones IR, Swai A, Taylor R, Miller M, Laker MF, Alberti KG. Lowering of plasma glucose concentrations with bezafibrate in patients with moderately controlled NIDDM. *Diabetes Care.* 1990;13:855-863.
77. Jones P, Kafonek S, Laurora I, Hunninghake D. Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (the CURVES study). *Am J Cardiol.* 1998;81:582-587.
78. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. *JAMA.* 1998;279:1615-1622.
79. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med.* 1995;333:1301-1307.
80. Karpf DB, Shapiro DR, Seeman E, et al. Prevention of nonvertebral fractures by alendronate: a meta-analysis. *JAMA.* 1997;277:1159-1164.
81. Black DM, Cummings SR, Karpf DB, et al, and the Fracture Intervention Trial Research Group. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet.* 1996;348:1535-1541.
82. Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women. *JAMA.* 1999;281:2189-2197.
83. Ettinger B, Black DM, Mitlak BH, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized trial. *JAMA.* 1999;282:637-645.