Surrogate End Points in Clinical Research: Hazardous to Your Health

David A. Grimes, MD, and Kenneth F. Schulz, PhD, MBA
Family Health International, Research Triangle Park, North Carolina

Surrogate end points in clinical research pose real danger. A surrogate end point is an outcome measure, commonly a laboratory test, that substitutes for a clinical event of true importance. Resistance to activated protein C, for example, has been used as a surrogate for venous thrombosis in women using oral contraceptives. Other examples of inappropriate surrogate end points in contraception include the postcoital test instead of pregnancy to evaluate new spermicides, breakage and slippage instead of pregnancy to evaluate condoms, and bone mineral density instead of fracture to assess the safety of depo-medroxyprogesterone acetate. None of these markers captures the effect of the treatment on the true outcome. A valid surrogate end point must both correlate with and accurately predict the outcome of interest. Although many surrogate markers correlate with an outcome, few have been shown to capture the effect of a treatment (for example, oral contraceptives) on the outcome (venous thrombosis). As a result, thousands of useless and misleading reports on surrogate end points litter the medical literature. New drugs have been shown to benefit a surrogate marker, but, paradoxically, triple the risk of death. Thousands of patients have died needlessly because of reliance on invalid surrogate markers. Researchers should avoid surrogate end points unless they have been validated; that requires at least one well done trial using both the surrogate and true outcome. The clinical maxim that “a difference to be a difference must make a difference” applies to research as well. Clinical research should focus on outcomes that matter. (Obstet Gynecol 2005;105:1114–8. © 2005 by The American College of Obstetricians and Gynecologists.)

Improper use of surrogate end points in clinical research carries real danger. This “quick and dirty” approach yields results faster than would use of important clinical outcomes, but the answers are often wrong. Regrettably, many researchers study surrogate outcomes that are not valid proxies for real events. Misleading results can lead to medical harm.1

WHAT IS A SURROGATE END POINT?
A surrogate end point is an outcome measure that substitutes for a clinical event of true importance. Also deemed “surrogate markers” or “intermediate measures,” these outcomes are commonly laboratory measurements or imaging studies thought to be involved in the causal pathway to a clinical event of interest. An example would be CD4 counts as a surrogate for survival of patients with acquired immunodeficiency syndrome (AIDS) or blood pressure as a surrogate for morbidity and mortality in patients taking antihypertensive medication.

PROS AND CONS
Surrogate markers have seductive appeal. Most importantly, they sometimes enable clinical studies to have smaller sample sizes and shorter follow-up periods than would otherwise be possible. Studying short-term changes in blood pressure is far easier than following women for years to study mortality rates. Alternatively, the clinical end point may be so invasive, uncomfortable, or expensive that surrogate outcomes are considered acceptable.2 On the other hand, surrogate end points may waste resources, provide ambiguous evidence, and not measure what one wants to study.3 Even if valid, surrogate end points may not be capable of extrapolation to other populations. Most worrisome, they frequently lead to patient harm.

CRITERIA FOR SURROGATE END POINTS
Criteria for a valid surrogate outcome have been proposed by several authors. Perhaps the most explicit is this: “a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true end point.”2

A risk factor is not necessarily a surrogate marker. Correlation with outcome is necessary but insufficient. As noted by Fleming and DeMets, “A correlate does not make a difference.”4 The surrogate must also fully capture the net effect of treatment on the outcome.4 Although many surrogate end points in the literature meet the first criterion (correlation with the clinical event), few
meet the more stringent second criterion (capture of treatment effect on the true outcome).

The proper development of surrogate markers requires conducting a trial with a given treatment and analyzing both the true and surrogate end points, for example, venous thrombosis and a coagulation test used as a proxy.5 Ironically, to establish surrogate end points, investigators must perform the very trial they had hoped to avoid. Hence, analysis of surrogate outcomes in a trial with true end points would be superfluous. If the surrogate proved adequate, however, future trials would benefit.

We will describe several reasonable surrogate markers and then describe serious (and sometimes deadly) errors stemming from the use of invalid surrogate end points. Finally, we provide examples of current concern in the contraception literature.

PROMISING SURROGATE END POINTS

In infertility treatment, a viable pregnancy at 20 weeks of gestation accurately predicts live birth. However, the need for such a surrogate is limited, because the true end point is readily measured with a few more months of observation.

Plasma human immunodeficiency virus (HIV) RNA is widely used as a surrogate marker in efficacy trials of HIV drug therapy. In confirmatory trials with 4 different drugs, this surrogate captured the treatment effect on clinical outcomes: reductions in disease progression and death.6 Even so, some argue that use of plasma HIV RNA as a single marker may be premature. “The purpose of a phase III clinical trial is to learn about how to treat the patient, not just the patient’s viral RNA!”6

Experimental and observational studies support the usefulness of adenomas (adenomatous polyps) as a surrogate marker for colorectal cancer.7 For example, a population-based study found that use of nonsteroidal anti-inflammatory drugs was associated with a reduced risk of both adenomas and colorectal cancer. Removing adenomas by colonoscopy is associated with a marked reduction in the risk of cancer. However, this surrogate has limitations as a cancer biomarker. Adenomas occur early in tumor development. Thus, adenomas might not capture the effect of an intervention that acted later in the process (for example, the transition from small to large adenoma or from large adenoma to cancer). The biological heterogeneity of adenomas might lead to incorrect inferences, because effects of an intervention might differ by histology.8,9

SERIOUS ERRORS

Ventricular arrhythmia is correlated with a nearly 4-fold increase in the risk of death from cardiac complications, particularly sudden death (Table 1). Hence, use of antiarrhythmic drugs should lower the risk. Three drugs (encainide, flecainide, and moricizine) nicely suppressed ventricular irritability and were approved by the U.S. Food and Drug Administration for treating severe arrhythmias. A large randomized controlled trial (Cardiac Arrhythmia Suppression Trial [CAST]) compared the 3 drugs with placebo in myocardial infarction patients having frequent premature ventricular contractions; sudden death was the primary outcome. These drugs decreased arrhythmias but paradoxically increased the risk of death from other causes when compared with placebo.10 Encainide and flecainide tripled the death rate. Nationwide, more than 200,000 seriously ill patients received these drugs, and thousands died needlessly as a result. The surrogate end point (arrhythmia) did not capture the effect of treatment on the true outcome (death).

Fluorides increase bone mineral density in women with osteoporosis. Hence, investigators presumed that sturdier bones would translate into fewer fractures. When a randomized controlled trial compared fluoride to placebo, the beneficial effect on bone mineral density was confirmed: in the lumbar spine, the increase was 35%. Paradoxically, both vertebral and nonvertebral fractures were more common among the women given fluoride. Their bones became more dense but more

<table>
<thead>
<tr>
<th>Condition</th>
<th>Treatment</th>
<th>Surrogate End Point</th>
<th>What Happened to Trial Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventricular arrhythmia10</td>
<td>Encainide, flecainide</td>
<td>Arrhythmia</td>
<td>Increased sudden death</td>
</tr>
<tr>
<td>Osteoporosis11</td>
<td>Fluorides</td>
<td>Bone mineral density</td>
<td>Increased fractures</td>
</tr>
<tr>
<td>Heart disease in postmenopausal women12</td>
<td>Conjugated equine estrogen and medroxyprogesterone acetate</td>
<td>Serum lipids</td>
<td>Increased heart disease</td>
</tr>
<tr>
<td>Heart disease13</td>
<td>Clofibrate</td>
<td>Exercise tolerance</td>
<td>Increased mortality</td>
</tr>
<tr>
<td>Heart failure14</td>
<td>Epoprostenol</td>
<td>Symptoms of heart failure</td>
<td>Increased mortality</td>
</tr>
<tr>
<td>Heart failure15</td>
<td>Ibopamine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
brittle, subject to more fractures. A meta-analysis of all the fluoride trials confirmed that with increasing doses of fluoride, the risk of nonvertebral fracture and gastrointestinal toxicity increases—without any benefit on vertebral fracture rates. The surrogate end point (bone mineral density) did not capture the effect of treatment on the true outcome (fracture).

Serum lipids correlate with the risk of cardiovascular disease. Some used this correlate to predict the effect of oral contraceptives on the cardiovascular health of women. For example, pills containing norethindrone have fewer adverse effects on lipids than do pills containing levonorgestrel. Hence, investigators inferred that the norethindrone pills were safer. However, when the large Royal College of Obstetricians and Gynaecologists’ oral contraception study examined myocardial infarction (the outcome of interest), no significant differences in risk emerged between the pill types (Croft P, Hannafor P. Risk factors for acute myocardial infarction in women [letter]. BMJ 1989;298:674). The surrogate end point (serum lipids) did not capture the effect of treatment on the true outcome (myocardial infarction).

Based on serum lipids, similar predictions were made for menopausal women. Many observational studies showed beneficial changes in serum lipids associated with estrogen therapy, and these were thought responsible for apparent reductions in cardiovascular mortality. However, when randomized controlled trials were finally done, the predicted improvements in lipids occurred but without the anticipated benefit in cardiac health. Moreover, in the Women’s Health Initiative trial, these beneficial lipid changes were associated with an increased risk of heart disease. The surrogate end point (serum lipids) did not capture the effect of treatment on the true outcome (heart disease).

Serum lipids have led physicians astray in other ways. Because elevated cholesterol levels are associated with coronary artery disease, lowering levels should reduce the risk of death. Thirty years ago, physicians commonly prescribed clofibrate to lower cholesterol levels when diet changes proved inadequate. When a randomized controlled trial was done to evaluate the effect of clofibrate on mortality, the results were shocking: a 17% increase in deaths compared with placebo. This increase in mortality persisted for 4 years after withdrawal of the drug. More than 5,000 asymptomatic persons died needlessly in the United States because of this treatment. The surrogate end point (serum cholesterol) did not capture the effect of treatment on the true outcome (death). Other drugs have had a paradoxical effect on mortality (Table 1).

CLOTTING CONUNDRUMS

In 1995, intense controversy erupted over alleged differences in risk of venous thrombosis and pulmonary embolism associated with different “generations” of oral contraceptives. The resulting “pill scare” caused an epidemic of unintended pregnancies and induced abortions. Appropriately, several large epidemiological studies have examined these rare events. More worrisome, however, has been the proliferation of laboratory studies focusing, not on venous thrombosis and pulmonary embolism, but on laboratory tests of no known relevance. Recent examples include sex hormone-binding globulin and resistance to activated protein C.

In the 1990s, sex hormone-binding globulin attracted renewed interest as a putative marker for the “androgenicity” of combined oral contraceptives. This concept posed several challenges. First, “androgenicity” is neither a word nor a clear, specific, and measurable outcome. Second, sex hormone-binding globulin is not a steroid hormone with an end-organ effect. Third, levels of sex hormone-binding globulin in pill users do not correlate with levels of free testosterone or with clinical outcomes of interest. Sex hormone-binding globulin has now come full circle, being proposed as a measure of “estrogenicity.” This neologism poses similar problems, both linguistic and endocrinological. As with “androgenicity,” “estrogenicity” is not a word nor a measurable outcome.

Two recent reports suggest that the problem of inappropriate surrogate end points is worsening. One study attempted to correlate risk of venous thrombosis with changes in sex hormone-binding globulin related to various oral contraceptives. The authors postulated that the inverse relationship envisioned between sex hormone-binding globulin and risk of clots might be a marker for “estrogenicity” and, thus, the risk of venous thrombosis. However, no clear link exists between sex hormone-binding globulin, “estrogenicity,” and thrombosis. As noted by the authors, “It would be preferable to have a surrogate end point for VTE [venous thromboembolism] that is more directly related to the hemostatic system.” The surrogate end point (sex hormone-binding globulin) did not capture the effect of treatment on the true outcome (venous thrombosis).

A small randomized controlled trial from the same center used another surrogate marker for thrombosis: resistance to activated protein C. The report correlated changes in sex hormone-binding globulin (a surrogate marker with no predictive value) with resistance to activated protein C (another surrogate marker with no predictive value). This presents a reductio ad absurdum.
Certain hereditary and acquired thrombophilias increase a woman’s risk of venous thrombosis. However, laboratory tests of coagulation processes (e.g., resistance to activated protein C) or serum-binding globulins (for example, ceruloplasmin or sex hormone–binding globulin) related to oral contraception have no predictive value. Four decades of research (and over 1,000 published articles in PubMed) have established that none predicts venous thrombosis. This dead horse re-relished articles in PubMed) have established that none predicts venous thrombosis. Four decades of research (and over 1,000 published articles in PubMed) have established that none predicts venous thrombosis.27–29 This dead horse requires no further flogging.

OTHER SURROGATE END POINTS IN CONTRACEPTION

Inappropriate use of surrogate end points extends beyond oral contraception. A common way of evaluating potential new vaginal spermicides is to perform postcoital tests after use of the product.30 This practice is irrational because the test is invalid31 and does not capture the effect of treatment on the true outcome (fertility).32

Slippage and breakage rates are frequent surrogate end points in studies of condoms.33 Slippage and breakage may be invalid surrogates for pregnancy, or alternatively, they may be valid surrogates that cannot be reliably measured. For example, in one trial nonlatex condoms had an 8-fold higher breakage rate than did latex condoms, yet pregnancy rates over 6 months of use were only modestly higher and not significantly different.34 Clearly, the key end point is pregnancy.35

Bone mineral density, another surrogate marker, recently prompted the U.S. Food and Drug Administration to issue a “black-box warning” in package labeling of depo-medroxyprogesterone acetate.36 Use for more than 2 years is now discouraged. Bone mineral density addresses bone quantity but not quality. Indeed, the genetics of bone mineral density and fracture appear to differ.37 Regrettably, no valid proxy for fracture is available.38 Depo-medroxyprogesterone acetate, a highly effective contraceptive, will now be used less because of worry about a surrogate marker (bone mineral density) that does not predict the true outcome (fracture).39

RECOMMENDATION

A finite amount of money and time can be spent on women’s health research. Every dollar spent studying surrogate markers of no clinical utility27–29,38 is a dollar not spent on worthwhile activities. Useless research diverts human and financial resources urgently needed elsewhere. Thus, such research has a net negative effect on women’s health.1

The medical literature is littered with thousands of useless or misleading studies of surrogate end points. Researchers should avoid surrogate end points in clinical trials unless they have been shown to be valid.4 This requires at least one well done clinical trial demonstrating that a surrogate end point captures the effect of the intervention on the outcome of interest (for example, death from all causes). Journals should be more critical of surrogate end points as well. The clinical maxim that “a difference to be a difference must make a difference” applies equally well to clinical research. Research should focus on outcomes that matter.

REFERENCES


Received December 2, 2004. Received in revised form January 10, 2005. Accepted January 13, 2005.