

Commentary

Debate: The slippery slope of surrogate outcomes

Ralph B D'Agostino Jr

Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA

Received: 25 September 2000

Accepted: 25 September 2000

Published: 11 October 2000

Curr Control Trials Cardiovasc Med 2000, 1:76–78

© Current Controlled Trials Ltd

(Print ISSN 1468-6708; Online 1468-6694)

Abstract

Surrogate outcomes are frequently used in cardiovascular disease research. A concern is that changes in surrogate markers may not reflect changes in disease outcomes. Two recent clinical trials (Heart and Estrogen/Progestin Replacement Study [HERS], and the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial [ALLHAT]) underscore this problem since their results contradicted what was expected based on the surrogate outcomes. The current regulatory policy to allow new therapies to be introduced onto the market based solely on surrogate outcomes may need to be reviewed.

Keywords: antihypertensive therapy, cardiovascular disease, clinical trials, surrogate outcomes

Introduction

There has been growing literature in recent years questioning the use of surrogate outcomes in medical research [3–5]. Two major randomized clinical trials, HERS [6] and ALLHAT [7], have underscored the fallacy of surrogate outcomes. The first question one must address is: What is a 'surrogate outcome'? A classic definition is provided by Temple [8]: "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."

However, drugs have multiple effects that are often not reflected by a surrogate measure, which is especially true for safety problems. In the HERS [6] study, for instance, the women receiving the active treatment (estrogen–progestin) had significantly more thromboembolic events than those patients on placebo (relative hazard [RH] = 2.89, 95% confidence interval [CI] = 1.50–5.58; $P=0.002$). This adverse drug effect is unrelated to the 'favorable' hormone effect on lipoprotein levels, generally thought of as a surrogate for a reduction in coronary events.

Let us consider some of the reasons why surrogates should or should not be employed in cardiovascular medicine. In a

perfect world, more information is clearly better than less. Understanding and measuring the true relationship between any therapy, on the one hand, and mechanistically and clinically meaningful endpoints, on the other, would thus be the natural goal of research. Measuring surrogate outcomes in addition to clinically meaningful endpoints would be quite useful in such studies. Having both surrogate outcomes and clinically meaningful disease outcomes measured on the same individuals may allow one to understand better the underlying pathophysiology.

Why are surrogates used?

So why would one suggest measuring the surrogate outcome and not the clinically meaningful endpoint? Some scientists have in recent years argued that the cost in both time and financial resources may be too great to conduct research that includes clinically meaningful endpoints. A trial to determine whether a new antihypertensive agent reduces the risk of mortality would take several years to perform and cost tens of millions of dollars, whereas one could determine whether the new drug lowers elevated blood pressure in several months at the fraction of the cost. This type of study may also lead to faster marketing since regulatory approval for many groups of drugs is currently based on surrogate outcomes. But is a study based on surrogate outcomes useful in today's medical climate? There are currently a host of antihypertensive drugs shown to lower blood pressure, and some of these improve survival. It is important to know which drugs provide optimal benefit to the patient. There are over 100 antihypertensive drugs on the US market alone today, and clinicians often do not know how these drugs compare in their ability to prevent disease outcomes. There is no reason to believe that hypertension is simply elevated blood pressure. Hypertension is much more complex. There is also no reason to believe that the non-blood pressure effects are clinically similar across the many classes of antihypertensive drugs.

One may thus ask why there is a need to introduce an additional antihypertensive drug without knowing the safety and long-term efficacy effects of the drug? For some diseases, such as cancer, where there are often no very effective therapies available to patients, there is a sense of urgency to develop and make new incompletely tested agents available. For cardiovascular disease, the sense of urgency is fortunately less pronounced.

The HERS trial

Consider the HERS [6] trial, a multicenter trial conducted in over 2700 women with coronary disease who were postmenopausal and younger than 80 years. The trial was conducted to determine whether a fixed dose of conjugated equine estrogen plus medroxyprogesterone acetate would decrease the rate of occurrence of non-fatal myocardial infarction or coronary heart disease (CHD)

death when compared with a placebo. Many investigators, prior to the start of the trial, clearly anticipated that the patients taking the active treatment would show improved outcomes. Previous epidemiological research had shown that women taking estrogen had lower low-density lipoprotein cholesterol and higher high-density lipoprotein cholesterol, and it was hypothesized that this hormone effect on these surrogate endpoints would result in a reduction in coronary events. The results of the HERS trial did not support this hypothesis [6]. Although the women receiving the active treatment demonstrated an 11% decrease in low-density lipoprotein and a 10% increase in high-density lipoprotein cholesterol, both statistically significant ($P < 0.001$), the trial found that the treatment and placebo groups had nearly identical risks for myocardial infarction (MI) or CHD death (RH = 0.99, 95% CI = 0.80–1.22). This trial illustrates how information based solely on the relationship between hormone therapy and the surrogate endpoints of low-density and high-density lipoprotein cholesterol did not correspond to an expected relationship between therapy and clinically meaningful endpoints. In addition, as stated earlier, the number of women in the active treatment arm with thromboembolic events was significantly higher than the number in the placebo group (RH = 2.89; $P = 0.002$). Estrogen may still have a positive effect on other clinically meaningful endpoints, but more research clearly ought to be carried out to clarify why some women appeared to respond unfavorably to hormone treatment.

The ALLHAT study

A second recent example comes from the ALLHAT study [7]. This study was a four-armed multicenter trial conducted on more than 44 000 patients, of which 15 268 patients were randomized to receive chlorthalidone and 9067 to receive doxazosin. The primary outcome of the study was fatal CHD or non-fatal MI, with secondary outcomes including all-cause mortality, stroke, and combined cardiovascular disease (CVD) (CHD death, non-fatal MI, stroke, angina, coronary revascularization, congestive heart failure [CHF] and peripheral artery disease). The difference in systolic blood pressure (SBP) between chlorthalidone and doxazosin in the ALLHAT study was 2–3 mmHg (with chlorthalidone patients having lower SBP), while there was no difference in diastolic blood pressure between the two treatments. The risk of CHF for the patients on doxazosin, however, was twice that of patients on chlorthalidone (relative risk = 2.04, 95% CI = 1.79–2.32). Based on data from the Systolic Hypertension in the Elderly Program [9,10], a 12 mmHg reduction in SBP was associated with a 49% reduction in CHF incidence. One could thus infer that the 2–3 mmHg difference in SBP in ALLHAT between treatments could possibly account for a 10–20% increase in CHF risk, but not a doubling of risk. These data suggest that chlorthalidone may have some favorable non-blood pressure effects, that

doxazosin may have some unfavorable non-blood pressure effects or a combination thereof. These results again demonstrate that measuring the effect of a drug on a surrogate endpoint, such as SBP, does not always provide adequate information concerning the impact of the drug on all clinically meaningful endpoints. Both surrogate and clinically meaningful endpoints were measured in ALLHAT, allowing information regarding the actions of the two drugs to be examined and compared. This would not have been the case had ALLHAT been conducted using only the surrogate endpoint of blood pressure.

Discussion

So where can surrogate endpoints be used? As already mentioned, surrogate endpoints are helpful in understanding the mechanism of action for different drugs. Most drugs generally do not 'prevent death' but rather influence biologic functions that are related to death. Surrogate endpoints are therefore very useful for advancing scientific knowledge during the drug development stages. Once a drug is developed, however, surrogate endpoints should not to be used for drug approval [3,5], particularly in the area of cardiovascular disease. The risks of introducing new incompletely tested drugs usually outweigh the potential benefits.

But now that we have argued this point, there may be a new opportunity for surrogate endpoints in future research, namely genetic studies. The underlying limitation with the use of surrogate endpoints is that the relationship between the surrogate endpoint and the clinically meaningful endpoint often is inconsistent or unpredictable. That is, for some patients lowering of elevated blood pressure may substantially reduce their risk for a subsequent MI (or other clinically meaningful endpoints) while for other patients similar lowering of blood pressure may have little effect in prognosis. Why this difference? Maybe this could be genetically determined. That is, there may exist certain genetic characteristics common to individuals with a strong causal relationship between blood pressure level and subsequent MI. If research could accurately identify and classify individuals into 'genetic classes', maybe surrogate endpoints within certain classes could be more useful in predicting reductions in disease endpoints. Our knowledge is not yet at this level of sophistication.

Conclusion

In conclusion, the reliance on surrogate endpoints to make drug approval and treatment decisions can be risky. A major limitation is safety. Even if we had a perfect surrogate for efficacy, it would only be helpful if we also had a perfect surrogate for safety. There are currently multiple treatment alternatives available to patients with most cardiovascular conditions, and the need to introduce new therapies without strong documentation is not clear. The pharmaceutical companies have an obligation to provide

physicians and patients with better information about clinical efficacy and safety of new products. The regulatory agencies ought to consider 'raising the bar' for drug approval to account for these considerations. The lessons concerning the fallacy of surrogate outcomes continue to accumulate. The HERS and ALLHAT trials are recent reminders that theories based on surrogate outcomes often do not translate to clinically meaningful outcomes.

Acknowledgement

This work was partially supported by NCI grant CA79934.

References

1. Buehler JW, Mulinare J: **Preventing neural tube defects.** *Pediatr Ann* 1997, **26**:535-539.
2. Ewan, PW: **Prevention of peanut allergy.** *The Lancet* 1998, **352**: 4-5.
3. Fleming TR, DeMets DL: **Surrogate end points in clinical trials: are we being misled?** *Ann Internal Med* 1996, **125**:605-613.
4. Sobel BE, Furberg CD: **Surrogates, semantics, and sensible public policy.** *Circulation* 1997, **95**:1661-1663.
5. Psaty BM, Weiss NS, Furberg CD, Koepsell TD, Siscovick DS, Rosendaal FR, Smith NL, Heckbert SR, Kaplan RC, Lin D, Fleming TR, Wagner EH: **Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease.** *JAMA* 1999, **282**:786-790.
6. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E: **Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women.** *JAMA* 1998, **280**:605-613.
7. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group: **Major cardiovascular events in hypertensive patients randomized to doxazosin vs. chlorthalidone.** *JAMA* 2000, **283**:1967-1975.
8. Temple RJ: **A regulatory authority's opinion about surrogate endpoints.** In *Clinical Measurement in Drug Evaluation*. Edited by Nimmo WS, Tucker GT. New York: Wiley; 1995.
9. Systolic Hypertension in Europe (Syst-Eur) Trial Investigators: **Randomized double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension.** *Lancet* 1997, **350**:757-764.
10. Kostis JB, Davis BR, Cutler JA, Grimm RH Jr, Berge KG, Cohen JD, Lacy CR, Perry HM Jr, Blafox MD, Wassertheil-Smoller S, Black HR, Schron E, Berkson DM, Curb JD, Smith WM, McDonald R, Applegate WB: **Prevention of heart failure by antihypertensive drug treatment in older persons with isolated systolic hypertension.** *JAMA* 1997, **278**:212-216.

Author's affiliation: Department of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA

Correspondence: Ralph B D'Agostino Jr, PhD, Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA.
Tel: +1 336 716 9410; fax: +1 336 716 5425;
e-mail: rdagosti@wfubmc.edu