

Surrogate end points: hopes and perils

Peer-reviewed author version

ALONSO ABAD, Ariel & MOLENBERGHS, Geert (2008) Surrogate end points: hopes and perils. In: Expert Review of Pharmacoeconomics & Outcomes Research, 8(3). p. 255-259.

DOI: 10.1586/14737167.8.3.255

Handle: <http://hdl.handle.net/1942/9635>

Surrogate endpoints: Hopes and Perils

Prof. Dr. Ariel Alonso¹ and Prof. Dr. Geert Molenberghs^{1,2}

¹Center for Statistics, Hasselt University Building D 3590 Diepenbeek, Belgium.

²Biostatistical Centre, Katholieke Universiteit Leuven, Belgium

ariel.alonso@uhasselt.be

fax: +32 (0)11 26 82 99

tel: +32 (0) 11 26 82 82

Abstract

In recent years, the cost of drug development has increased the demands on efficiency in the selection of suitable drug candidates; surrogate endpoints have emerged to improve this process, hoping that they can help reducing duration and cost of clinical trials. They can, additionally, help solving ethical issues when measuring the clinical endpoint involves the application of risky or uncomfortable medical procedures. However, the very mention of surrogate endpoints has always been controversial, owing in part to unfortunate historical events. As a consequence, there is growing consensus on the use of validated surrogates only. Here, we discuss some of the validation strategies that have recently been proposed and consider the future of surrogate endpoints in clinical research.

Keywords: biomarkers, surrogate marker, meta-analysis, information theory.

1. Surrogate Endpoints: Motivations and Antecedents

The rising costs of drug development and the challenges of new and re-emerging diseases are putting considerable demands on efficiency in the drug candidates selection process. A very important factor influencing duration and complexity of this process is the choice of endpoint used to assess drug efficacy. Often, the most sensitive and relevant clinical endpoint might be difficult to use in a trial. This happens if measurement of the clinical endpoint (1) is costly (e.g., to diagnose “cachexia”, a condition associated with malnutrition and involving loss of muscle and fat tissue, expensive equipment measuring content of nitrogen, potassium and water in patient’s body is required); (2) is difficult (e.g.,

involving compound measures such as encountered in quality-of-life or pain assessment); (3) requires a long follow-up time (e.g., survival in early stage cancers); or (4) requires a large sample size because of low event incidence (e.g., short-term mortality in patients with suspected acute myocardial infarction). An effective strategy is then proper selection and application of biomarkers for efficacy, replacing the clinical endpoint by a biomarker that is measured more cheaply, more conveniently, more frequently, or earlier. From a regulatory perspective, a biomarker is considered acceptable for efficacy determination only after its establishment as a valid indicator of clinical benefit, i.e., after its validation as a surrogate marker[1].

These considerations naturally lead to the need of proper definitions. An important step came from the Biomarker Definitions Working Group[2], their definitions nowadays widely accepted and adopted. A clinical endpoint is considered the most credible indicator of drug response and defined "a characteristic or variable that reflects how a patient feels, functions, or survives." During clinical trials, endpoints should be used, unless a biomarker is available that has risen to the status of surrogate endpoint. A biomarker is defined as a characteristic that can be objectively measured as an indicator of healthy or pathological biological processes, or pharmacological responses to therapeutic intervention. A surrogate endpoint is a biomarker, intended for substituting a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit, harm, or lack of these.

Surrogate endpoints have been used in medical research for a long time[3,4]. Owing to unfortunate historical events and in spite of potential advantages, their use has been surrounded by controversy. The best known case is the approval by the Food and Drug Administration (FDA) of three antiarrhythmic drugs: encainide, flecainide, and moricizine. The drugs were approved because of their capacity to effectively suppress arrhythmias. It was believed that, because arrhythmia is associated with an almost fourfold increase in the rate of cardiac-complication-related death, the drugs would reduce the death rate. However, a post-marketing trial showed that the active-treatment death rate was double the placebo rate. A risk was also detected for moricizine[5]. Another example came with the surge of the AIDS epidemic. The impressive early therapeutic results obtained with zidovudine, and the pressure for

accelerated evaluation of new therapies, led to the use of CD4 blood count as a surrogate endpoint for time to clinical events and overall survival[6], in spite of concern about its limitations as a surrogate marker for clinically relevant endpoints[7].

The main reason behind failures was the incorrect perception that surrogacy simply follows from the association between a potential surrogate endpoint and the corresponding clinical endpoint, the mere existence of which is insufficient for surrogacy [4]. Even though the existence of an association between the potential surrogate and the clinical endpoint is undoubtedly a desirable property, what is required to replace the clinical endpoint by the surrogate is that the effect of the treatment on the surrogate endpoint reliably predicts the effect on the clinical endpoint. Partly owing to the lack of appropriate methodology, this condition was not checked in the early attempts and, consequently, negative opinions about the use of surrogates in the evaluation of treatment efficacy emerged[4,8,9].

2. Why Surrogate Endpoints are Still Being Considered

Currently, the steady advance in many medical and biological fields is dramatically increasing the number of biomarkers and hence potential surrogate endpoints. Additionally, an increasing number of new drugs have well-defined mechanisms of action at molecular level, allowing drug developers to measure the effect of these drugs on the relevant biomarkers[10]. There is also increasing public pressure for fast approval of promising drugs, which will have to be based on biomarkers rather than on long-term, costly clinical endpoints[11]. Obviously, the pressure will be especially high when a rapidly increasing incidence of the targeted disease could become a serious threat to public health or the patient's (quality of) life.

Shortening the duration of clinical trials not only can decrease the cost of the evaluation process but also limit potential problems with noncompliance and missing data, which are more likely in longer studies [1,14].

Surrogate endpoints can play a role in the earlier detection of safety signals that could point to toxic problems with new drugs. The duration and sample size of clinical trials aimed at evaluating the

therapeutic efficacy of new drugs are often insufficient to detect rare or late adverse effects[12,13]; using surrogate endpoints in this context might allow one to obtain information about such effects even during the clinical testing phase.

Discoveries in medicine and biology are further creating a exciting range of possibilities for the development of potentially effective treatments. This is an achievement, but it also faces us with the challenge of coping with a large number of new promising treatments that should be rapidly evaluated. This is already clear in oncology, because the increased knowledge about the genetic mechanisms operating in cancer cells led to proposing of novel cancer therapies, such as the use of a genetically-modified virus that selectively attacks p53-deficient cells, sparing normal cells[15]. Validated surrogate endpoints can offer an efficient route.

The role of surrogate endpoints may depend on the trial's phase. Nowadays, their use is more accepted in early phases of clinical research, such as in phase II or early phase III clinical trials. Using them to substitute for the clinical endpoint in pivotal phase III trials or to replace the clinical endpoint altogether in all clinical research past a certain point is, however, a topic of ongoing debate. It is difficult to precisely define the future role of surrogate endpoint in the various trial phases. Ultimately, the combination of medical and statistical elements, together with practical and economical considerations, will help answer this question.

While the huge potential of surrogate endpoints to accelerate and improve the quality of clinical trials is unquestioned, the above considerations indicate that only thoroughly evaluated surrogates should be used, a point taken up next.

3. Statistical Evaluation of Surrogate Endpoints

While the past failures have led some to conclude that surrogate endpoints should be avoided altogether, practice has shown that sometimes they are the only route to evaluate a new drug, provided their validity should be carefully evaluated, using a principled statistical framework, prior to use. Such methods have become the subject of intensive research. Like in most clinical decisions, statistical

arguments and clinical and biological evidence need to be juxtaposed.

The first formal framework dates back to 1989 when Prentice proposed a definition of surrogate endpoints and outlined criteria, all within a hypothesis testing paradigm[16]. Much debate ensued, for the criteria set out by Prentice are not straightforward to verify[17,18] and are only equivalent to the definition for binary endpoints[1,19]. Freedman supplemented Prentice's approach by introducing the proportion of treatment explained (PTE), aimed at measuring the proportion of the treatment effect mediated by the surrogate[18]. This proposal was important in that it shifted the interest in the validation of surrogate endpoints from significance testing to estimation. However, it is itself surrounded with difficulties[1]. Buyse and Molenberghs showed that PTE can be decomposed in three different quantities: a scale parameter, the relative effect, and the adjusted association. The relative effect is the ratio of the effects of treatment upon the true and the surrogate endpoint; the treatment-adjusted association is the correlation between the surrogate and the clinical endpoint after adjusting by treatment. This decomposition is reminiscent of the two dimensions of the problem: the first one relates to capability of the surrogate to predict the treatment effect on the clinical endpoint; the second one describes its capability to predict the outcome of the clinical endpoint[1,19].

All earlier proposals were based on data coming from a single trial and therefore lacked treatment-effect replication. Within this framework, this limitation is circumvented through implicit, unverifiable assumptions. Trying to solve this problem, Daniels and Hughes, and Buyse and colleagues shifted to a meta-analytic context, additionally allowing for trial-level replication. These authors focused on continuous responses, employed linear mixed-effects models[15,20,21], and quantified the quality of a surrogate endpoint using two coefficients of determinations: R_{ind}^2 and R_{trial}^2 .

The R_{trial}^2 measures how precise the effect of treatment on the clinical endpoint can be predicted, using the treatment effect on the surrogate. It is unitless and ranges in the unit interval. If $R_{trial}^2 = 1$, then the treatment effect on the clinical endpoint can be predicted without error using the treatment effect on the surrogate, whereas $R_{trial}^2 = 0$ implies that both treatment effects are independent and no meaningful prediction is possible. The R_{ind}^2 is interpreted similarly, quantifying how precise the outcome on the

clinical endpoint can be predicted using the outcome on the surrogate. One can also derive prediction equations for the true treatment effect in a new trial, based on the treatment effect of the surrogate endpoint and similar equations can also be derived to predict the outcome on the clinical endpoint using the observed outcome on the surrogate.

This meta-analytic method fully captures both dimensions of the surrogacy problem and agrees with the Biomarker Definitions Working Group's definitions. Nevertheless, an obvious question is which of these two dimensions is the most important one in practice. There is no single answer. For instance, for a trialist who only wants to use the surrogate to predict the treatment effect on the clinical endpoint, the trial dimension is the more relevant one, with the reverse holding for the treating physician who has observed a tumor response in a patient and wants to know how this can predict the survival of the said patient.

Many extensions of this meta-analytic approach have been introduced to evaluate surrogacy when the candidate surrogate and the clinical endpoint are not continuous or are of a different nature like, for instance, when both the surrogate and the clinical endpoint are binary[22], when one of them is a time-to-event and the other one is categorical[1,23,24], when both endpoints are repeatedly measured over time[25], etc. Unfortunately, each generalization led to different surrogacy quantification, which could potentially lead to different interpretations.

Alonso and Molenberghs[26] proposed a unifying framework for the evaluation of surrogate endpoints using information theory. While the measures, coming from the framework of Buyse and colleagues, do not readily generalize to settings with non-normal and/or multivariate outcomes, the information-theoretic approach applies to a wide variety of settings (normal, binary, categorical, and longitudinal outcomes), with the quantities previously introduced in the literature following as special cases. The scattered set of proposals made earlier thus gets a theoretical basis.

Recently, a new approach has been introduced based on causal-inference concepts[27,28]. This method somehow returns to the original idea of Prentice and aims at studying the position of the surrogate within the causal path leading from the treatment to the clinical endpoint. Even though this

approach reverts to the single-trial setting and therefore strong assumptions are imperative, it can certainly be a promising research line, especially to evaluate surrogacy in initial stages when little information about the surrogate is available.

All these examples show that the development of methodological tools to evaluate surrogate endpoints is currently an active field of research and likely more methods and extensions are going to be proposed in the near future.

4. Expert Commentary

The initial naive enthusiasm that accompanied the first use of surrogate endpoints and the deception and skepticism that followed after the first failures are now being substituted by a more scientific and objective comprehension of their potentials and limitations. It is now clear that surrogate endpoints are a powerful tool that can play an important role in the development and evaluation of certain types of drugs. It is also clear that they need to be properly evaluated. Numerous statistical methods have been proposed in recent years to approach this validation problem and the work continues.

At the same time, regulatory agencies around the globe, have developed new policies and methods to accelerate the approval of certain types of drugs through the use of surrogate endpoints. In the United States, accelerated approval, sometimes referred to as “conditional approval” or subpart H, refers to an acceleration of the overall development plan by allowing submission of an application and, if approved, marketing of a drug based on the evidence obtained, for instance, using a surrogate endpoint, with further studies demonstrating that direct patient benefit is underway. In the same way, the European regulatory agency has developed a set of regulations that are converging to an accelerated approval system like in the United States, perhaps with more flexibility[1].

Manifestly, the new scientific developments in the validation of surrogate endpoints together with the clear regulations established by leading regulatory agencies in the world will allow, in the near future, a rational and efficient use of surrogate endpoints that will increase their potential benefits, thereby minimizing, the possible risks associated with their use.

5. Key Issues

- The high costs of drug development and the challenge implied by new and re-emerging diseases are putting considerable pressure on the development of efficient methods to rapidly evaluate suitable drug candidates.
- Surrogate endpoints can considerably accelerate the clinical evaluation of new drugs, reducing the cost and increasing the reliability of our studies.
- Surrogate endpoints that have not been properly validated can produce misleading results.
- Recently, a number of statistical methods have been proposed to evaluate surrogate endpoints.
- Regulatory agencies in the United States and Europe have put in place a set of methods and mechanism that regulate and formalize the use of surrogate endpoints in clinical research.

6. References

[1] * * * Burzykowski T, Molenberghs G, and Buyse, M. The Evaluation of Surrogate Endpoints. New York: Springer-Verlag (2005).

A nice compendium giving the stage of the art in surrogate marker validation

[2] * * * Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapy*, 69, 89—95 (2001).

A nice compendium including the most widely accepted definitions of clinical endpoint, surrogate endpoint and biomarker.

[3] Ellenberg SS and Hamilton JM. Surrogate endpoints in clinical trials:cancer. *Statistics in Medicine*, 8, 405–413 (1989).

[4] Fleming TR and DeMets DL. Surrogate endpoints in clinical trials:are we being misled? *Annals of Internal Medicine*, 125, 605–613 (1996).

[5] Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary Report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infraction. *New England Journal of Medicine*, 321, 406–412 (1989).

[6] De Gruttola V and Tu XM. Modelling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, 50, 1003–1014 (1995).

[7] Lagakos SW and Hoth DF. Surrogate markers in AIDS: Where are we? Where are we going? *Annals of Internal Medicine*, 116, 599–601 (1992).

[8] De Gruttola V, Fleming TR, Lin DY, and Coombs R. Validating surrogate markers -are we being naive? *Journal of Infectious Diseases*, 175, 237–246 (1997).

[9] Fleming TR. Surrogate endpoints in clinical trials. *Drug Information Journal*, 30, 545–551 (1996).

[10] Ferentz AE. Integrating pharmacogenomics into drug development. *Pharmacogenomics*, 3, 453–467 (2002).

[11] Lesko LJ and Atkinson AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annual Review of Pharmacological*

Toxicology, 41, 347–366 (2001).

[12] Dunn N and Mann RD. Prescription-event and other forms of epidemiological monitoring of side-effects in the UK. *Clinical and Experimental Allergy*, 29, 217–239 (1999).

[13] Jones TC. Call for a new approach to the process of clinical trials and drug registration. *British Medical Journal*, 322, 920–923 (2001).

[14] Verbeke G and Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag (2000).

[15] Heise C, Sampson-Johannes A, Williams A, McCormick F, Von Hoff DD, and Kirn DH. ONYX015, an E1B gene-attenuated adenovirus, causes tumor-specific cytolysis and antitumoral efficacy that can be augmented by standard chemo-therapeutic agents. *Nature Medicine*, 3, 639–645 (1997).

[16] Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. *Statistics in Medicine*, 8, 431–440 (1989).

First formal definition of surrogacy and operational criteria.

[17] Fleming TR, Prentice RL, Pepe MS, and Glidden D. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine*, 13, 955–968 (1994).

[18] Freedman LS, Graubard BI, and Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, 11, 167–178 (1992).

[19] Buyse M and Molenberghs G. The validation of surrogate endpoints in randomized experiments. *Biometrics*, 54, 1014–1029 (1998).

[20] Buyse M, Molenberghs G, Burzykowski T, Renard D, and Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, 1, 49–67 (2000a).

A complete introduction to the meta-analytic approach.

[21] Daniels MJ and Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine*, 16, 1515–1527 (1997).

[22] Renard D, Geys H, Molenberghs G, Burzykowski T, and Buyse M. Validation of surrogate

endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal*, 44, 921–935 (2002).

[23] Burzykowski T, Molenberghs G, Buyse M, Geys H, and Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. *Applied Statistics*, 50, 405–422 (2001).

[24] Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, and Piedbois P, for the Meta-Analysis Group In Cancer . Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. *Lancet*, 356, 373–378 (2000b).

[25] Alonso A, Geys H, Molenberghs G, and Kenward MG. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics*, 60, 845–853 (2004).

[26]. . . Alonso A and Molenberghs G. Surrogate marker evaluation from an information theoretic perspective. *Biometrics*, 63, 180–186 (2007).

A full description of the use of information theory within the meta-analytic framework.

[27] Chen H, Geng Z, and Jia J. Criteria for surrogate end points. *Journal of the Royal Statistical Society B*, 69, 919–932 (2007).

[28]. . . Frangakis C E and Rubin DB. Principal stratification in causal inference. *Biometrics*, 58, 21-29 (2002).

A validation approach based on causal inference.