Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective

DOI: 10.1177/0962280207081851
Handle: http://hdl.handle.net/1942/8744

Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective

Ariel Alonso and Geert Molenberghs
Center for Statistics, Hasselt University, Diepenbeek, Belgium
Email: ariel.alonso@uhasselt.be

Abstract

The last two decades have seen a lot of development in the area of surrogate marker validation. One of these approaches places the evaluation in a meta-analytic framework, leading to definitions in terms of trial- and individual-level association (Buyse *et al.* 2000). A drawback of this methodology is that different settings have led to different measures at the individual level. Using information theory, Alonso *et al.* (2006) proposed a unified framework, leading to a new definition of surrogacy, that offers interpretational advantages, and is applicable in a wide range of situations. In the present work, we illustrate how this information-theoretic approach can be used to evaluate surrogacy when both endpoints are of a time-to-event type. Two meta-analyses, in early and advanced colon cancer, respectively, are then used to evaluate the performance of time to cancer recurrence as a surrogate for overall survival.

Keywords: Information theory; Surrogate endpoints, Survival analysis.

1 Introduction

Information theory is a relatively new branch of the mathematical theory of probability and statistics, made mathematically rigorous only from 1940s onwards. The term *information theory* does not have a unique definition. Broadly speaking, information theory deals with the study of problems concerning any system. In fact, it has been applied in a variety of fields and plays a prominent role in modern

communication theory, which formulates a communication system as a stochastic or random process.

The theory has its mathematical roots connected with the idea of disorder or entropy used in thermodynamics and statistical mechanics. Some of the first attempts for formalizing these information-theoretic ideas were undertaken by Nyquist (1924) and by Hartley (1928), who recognized the logarithmic nature of the measure of information. A major contribution in this area came in 1948 when Shannon published a remarkable paper on the properties of information sources and the communication channels used to transmit the output of these sources.

The fundamental quantities of information theory, entropy, relative entropy, and mutual information, are defined as functionals of probability distributions and can be placed within a probabilistic framework. R. A. Fisher's (1925) measure of the amount of information supplied by data about an unknown parameter is well known to statisticians. This measure is the first use of information in mathematical statistics and was introduced especially for the theory of statistical estimation. A quarter of a century later, Kullback and Leibler (1951) studied another measure of information from a statistical point of view, involving two probability distributions associated with the same experiment. Other proposals to measure information have appeared in the literature over the last 20 years.

Alonso *et al.* (2006) used information-theoretic ideas to introduce a new and simple definition of surrogacy that possesses an appealing interpretation. This link between information theory and surrogate marker evaluation allows us to approach the validation problem in a unified way when the true and surrogate endpoints are of a different nature. In the present work, we illustrate how this information-theoretic approach can be applied when both the surrogate and the true endpoint are of a

time-to-event type.

Section 2 summarizes some of the main developments that have appeared in the surrogate marker literature over the last twenty years. In Section 3, we introduce the information-theoretic approach to surrogate marker evaluation. In Section 4, the methodology presented in Section 3 is applied to evaluate time to cancer recurrence as a surrogate marker for overall survival in early and advanced colon cancer.

2 Surrogate Marker Validation

The endpoint chosen to evaluate the efficacy of a new treatment is one of the most important factors influencing the complexity and duration of modern clinical trials. Frequently, the most sensible and relevant clinical endpoint, the so-called "true" endpoint (T), is difficult to use in a clinical trial, for example, when its use would imply risky manipulations of the patient or would increase the duration and/or cost of the study. In such situations, an attractive and sensible solution is to replace the "problematic" true endpoint by another one that can be measured more conveniently, a so-called "surrogate" endpoint (S).

In a seminal paper, Prentice (1989) provided a definition and a set of criteria that have formed the basis for a lot of subsequent work. Freedman *et al* (1992) introduced the *proportion of treatment explained* to quantify how much of the treatment effect on the true endpoint is captured by the surrogate endpoint. Buyse *et al* (1998) decomposed the proportion of treatment explained into the *relative effect* and the *adjusted association*, and argued in favor of these quantities instead. These proposals were formulated assuming that the validation of a surrogate is based on data from a single randomized clinical trial. This, however, leads to problems with untestable assumptions and too low a statistical power. To overcome these problems, Albert *et*

al (1998) suggested to combine information from several groups of patients, such as, for example, multi-center trials or meta-analyses. Approaches following these ideas were implemented by Daniels *et al* (1997), Gail *et al* (2000), and Buyse *et al* (2000). The latter suggested a multi-trial approach that led to a new definition of surrogacy in terms of the quality of both trial-level and individual-level association between the surrogate and the true endpoint. In their approach, the quality of a surrogate at the trial level is assessed by means of a coefficient of determination R^2_{trial}. At the individual level, the squared correlation R^2_{ind} between the surrogate and true endpoint, after adjustment for both the trial effects and the treatment effects, is used.

In this meta-analytic scenario, several individual-level measures have been proposed. In the binary-binary setting, Renard *et al.* (2002) used the correlation between two latent variables $R^2_{\text{ind}} = \text{corr}(\tilde{S},\tilde{T})$ to define individual-level surrogacy and alternatively defined $R^2_{\text{ind}} = \psi$, the global odds ratio between both endpoints estimated from a bivariate Plackett-Dale model. When the true endpoint is a survival time and the surrogate is a longitudinal sequence, Renard *et al* (2003), using Henderson's model, proposed to study the individual level based on a time function defined as $R^2_{\text{ind}}(t) = \text{corr}[W_1(t), W_2(t)]^2$, where $[W_1(t), W_2(t)]$ is a latent bivariate Gaussian process. When both responses are measured longitudinally, the so-called *variance reduction factor* (VRF), a canonical-correlation based quantity θ_p, and R^2_Λ have been proposed to evaluate surrogacy (Alonso *et al.* 2006, Burzykowski, Molenberghs and Buyse 2005). Additionally, the VRF, θ_p and R^2_Λ can be incorporated into a more general framework allowing for interpretation in terms of canonical correlations of the error vectors, based on which these authors defined different families of individual-level parameters. Other proposals have been suggested in other settings.

All of these examples clearly show one of the main limitations of the meta-analytic

methodology so far: different settings require different definitions. For some of these settings, researchers have proposed to estimate the association between both endpoints at a certain latent level, which, while mathematically convenient, can be clinically less relevant or difficult to interpret. Moreover, in all of the previous cases a joint model for both endpoints needs to be fitted. This can represent a very serious computational burden in many practical situations and, in addition, most of these models are not implemented in standard software packages rendering the methodology difficult to apply.

To overcome these limitations, Alonso *et al.* (2006) used information theory to create a unified framework, leading to a definition of surrogacy with an intuitive interpretation and applicable in a wide range of situations. Their approach also enhances insight into the chances of finding a good surrogate endpoint in a given situation. They further showed that some of the previous proposals follow as special cases of this information-theoretic approach. In the following section, we outline this methodology.

3 Information-theoretic Approach

Alonso *et al.* (2006) propose to term S a good surrogate for T at the individual level if uncertainty about T is reduced by a "large" amount when S is known; the corresponding definition for the trial level is that a good surrogate implies that the uncertainty about the effect of treatment on T is reduced by knowledge about the effect of treatment on S. These definitions, in spite of being based on formal concepts rooted in information theory, are simple and intuitive. Note that the general idea behind surrogacy is to reduce the uncertainty, or equivalently, to gain information about a "problematic" true endpoint through the use of a surrogate. At the trial level the situation is similar: we want to gain information about the unobserved

treatment effect on the true endpoint using the treatment effect on the surrogate.

To quantify the proportion of the uncertainty about the true endpoint that the surrogate can explain, these authors proposed to use the so-called R_h^2, defined as:

$$R_h^2 = \frac{\mathrm{EP}(T|Z) - \mathrm{EP}(T|Z,S)}{\mathrm{EP}(T|Z)}.$$

where $\mathrm{EP}(X) = \frac{1}{(2\pi e)^n} e^{2h(X)}$ denotes the so-called power entropy of the random variable X with density function f and h denotes its entropy defined as $h(X) = E[-\log f(X)]$. Note that Z represents treatment allocation. When the conditional distribution of T (and/or S), given Z differs substantially from the marginal distribution of T (and/or S), it follows that a substantial portion of the total variability in the outcome is explained by treatment.

R_h^2 satisfies a number of useful properties: (i) $0 \leq R_h^2 \leq 1$; (ii) $R_h^2 = 0$ if and only if (T, S) are independent; (iii) R_h^2 is symmetric in (T, S); (iv) R_h^2 is invariant under bijective transformations of T and S, in the sense that there is a 'one-to-one onto' mapping between S and T; (v) When $R_h^2 \longrightarrow 1$ for continuous models, there is usually some degeneracy appearing in the distribution of (T, S), i.e., often $T = \phi(S)$ for some nontrivial function ϕ. The latter means that then there exists a deterministic relationship between T and S. In a meta-analytic framework with N clinical trials, one could have different, trial-specific R_{hi}^2. In this setting, a plausible approach is to use a meta-analytic R_h^2 defined as $R_h^2 = \sum_{i=1}^N \alpha_i R_{hi}^2$, where $\alpha_i > 0$ for all i and $\sum_{i=1}^N \alpha_i = 1$. The α_i's could be chosen to represent (un)weighted averages of the trial-specific individual-level surrogacies R_{hi}^2, to produce an overall individual-level surrogacy. Many choices for the α_i's are possible, giving rise to a family of measures. Clearly, these calculations require data on S and T to be available from all trials. Similar families have been proposed to evaluate individual level surrogacy in other settings (Burzykowski, Molenberghs, and Buyse 2005). In

the following section, we will apply these ideas to analyze the two case studies.

4 Analysis of the Case Studies

In the present section we will use the meta-analytic framework described in section 2 to evaluate the performance of time to cancer recurrence as surrogate marker for overall survival using data from two meta-analyses in early colon cancer (Sargent 2005) and advanced colon cancer (MAGIC 2004).

This meta-analytic approach identifies two dimensions in the surrogate marker problem, i.e., the trial and individual dimension. The information-theoretic measure R_h^2 can be used to measure either the individual- or trial-level surrogacy (depending on the context); the calculations at each of the levels are similar but different. This approach has advantages over previously introduced measures, such as the R^2_{trial} and R^2_{ind} introduced by Buyse *et al.* (2000), because it can be applied to a wide variety of data types.

We will therefore approach the problem using the information-theoretic methodology introduce in Section 3 and use the R_h^2 to quantify the individual-level and trial-level surrogacies. It is important to point out that, when applied at the trial level, and assuming a linear functional relationship between the pairs of trial-specific treatment effects on the true and surrogate endpoints, respectively, the R_h^2 equals the R^2_{trial}. This equality allows us to give a new interpretation to the R^2_{trial}. Indeed, the R^2_{trial} can now be interpreted as the proportion of all the uncertainty about the treatment effect on the true endpoint that will be explained by the treatment effect on the surrogate. When a more complex functional form is necessary to describe the relationship between both treatment effects at the trial level, the R^2_{trial} becomes inapplicable, given the fact that the linear mixed model behind the calculations

might be overly simple, and hence the R_h^2 is a viable alternative to quantify surrogacy at the trial level as well. This illustrates that the surrogacy measures proposed by Buyse *et al.* (2000) can be seen as special cases of the more general framework based on the information-theoretic approach previously presented.

4.1 Advanced Colon Cancer

The analysis was based on data coming from 10 clinical trials in advanced colon cancer (MAGIC 2004). To evaluate trial level surrogacy two different approaches were used. In the first approach, two independent proportional hazard models were fitted at the first stage within each trial for the surrogate and the true endpoint, respectively. These models only included the treatment variable indicator, Z, as a covariate. In the second stage, the maximum likelihood estimates of the trial specific treatment effects on true endpoint ($\hat{\beta}_i$) and the surrogate ($\hat{\alpha}_i$) were used to estimate the R^2_{trial}, which is the same as the version of the R_h^2 used at the trial level in this case. The latter fact has been established in Alonso and Molenberghs (2006).

In the second approach, the association between both endpoints was taken into account by fitting a shared gamma frailty model within each trial at the first stage. The previous procedure is equivalent to using a Clayton copula with margins modelled using a proportional hazard regression. Here again, in the second stage, the maximum likelihood estimates of the trial-specific treatment effects were used to quantify the trial-level surrogacy.

To evaluate the individual-level surrogacy, we first defined a time-dependent covariate $S(t)$ which takes value 0 until the surrogate endpoint occurs and 1 thereafter. The following two models were fitted:

$$h_{ij}(t) \;=\; h_{i0}(t)\exp\{\beta_i Z_{ij}\},$$

$$h_{ij}(t) \quad = \quad h_{i0}(t)\exp\{\beta_{Si}Z_{ij} + \gamma_i S_{ij}(t)\},$$

where i denotes the trial and j denotes the subject. Using these two models, Alonso *et al.* (2006) showed that under some general conditions R_h^2 can be estimated using the so-called likelihood reduction factor (LRF) introduced in Alonso *et al.* (2004).

Table 1 displays the results for both the trial- and the individual-level surrogacy. At the trial level some convergency problems were encountered, when fitting the Clayton copula model to data from one particular trial and therefore the results shown in the table are calculated excluding this trial. Both approaches used to quantify trial-level surrogacy, i.e., using separate models on the one hand and the Clayton copula on the other hand, lead to similar point estimates. These point estimates hint on the presence of a large association at the trial level. However, the wide confidence intervals obtained in both cases do not rule out a weaker association.

When the problematic trial was taken into account, the approach using independent Cox models produced a $\hat{R}^2_{\text{trial}} = 0.82$ ($CI = [0.40; 0.95]$) and at the individual level $\hat{R}^2_h = 0.84$ ($CI = [0.82; 0.85]$). Clearly, the inclusion of this trial seems to have an important impact on the individual-level surrogacy while less so at the trial level. A closer exploration of the trial producing convergency problems showed that in this study the time between cancer recurrence and death was considerably smaller. However, whether this trial is included or not we always observed a large value of R_h^2 indicating that the surrogate can explain a large proportion (more than 76%) of our uncertainty about the true endpoint.

4.2 Early Colon Cancer

This meta-analysis contains data coming from more than ten thousands patients included in 10 early colon cancer trials Sargent *et al.* (2006).

Table 1: Advanced Colon Cancer excluding the trial generating convergency problems. Trial-level and individual-level measures of surrogacy.

Parameter	Estimate [95% C.I.]
Trial-level measures	
$\hat{R}^2_h \equiv \hat{R}^2_{\text{trial}}$ (separate models)	0.82 [0.41; 0.96]
$\hat{R}^2_{\text{trial}}$ (Clayton copula)	0.88 [0.52; 0.97]
Individual-level measures	
$\hat{R}^2_h$	0.76 [0.74; 0.78]
Percent censored	21%

Table 2: Early Colon Cancer. Trial-level and individual-level surrogacy.

Parameter	Estimate [95% C.I.]
Trial-level measures	
$\hat{R}^2_h \equiv \hat{R}^2_{\text{trial}}$ (separate models)	0.85 [0.53; 0.96]
$\hat{R}^2_{\text{trial}}$ (Clayton copula)	0.82 [0.44; 0.95]
Individual-level measures	
$\hat{R}^2_h$	0.84 [0.83; 0.85]
Percent censored	55%

Like in the previous case study, the analysis was again based on data coming from 10 clinical trials in early colon cancer. The same approaches used in the previous example were applied to evaluate the trial- and individual-level surrogacy. Table 2 summarizes the results. Here again, a very strong association was observed at the individual level. The large value obtained for the R^2_h indicates that time to cancer recurrence can explain more than 84% of the uncertainty about the survival of the patient. Once again, large point estimates for R^2_{trial} were observed, notwithstanding the wide confidence intervals hamper our interpretation of these point estimates.

5 Discussion

Based on a meta-analytic paradigm, Alonso *et al.* (2006) introduced an information-theoretic approach to evaluate surrogacy. This approach leads to a simple yet meaningful definition of surrogacy and offers a unified approach to surrogate marker evaluation.

While the R^2 measures, coming from the framework of Buyse *et al.* (2000), do not readily generalize to settings with non-normal outcomes, the R_h^2 applies to a wide variety of settings (normal, binary, categorical, and longitudinal outcomes) and reduces, in all of these specific settings, to the quantities previously introduced in the literature. This provides a theoretical basis for the scattered set of proposals made earlier in the literature.

In the present work, we have used this information-theoretic approach to evaluate the performance of time to cancer recurrence as surrogate marker for overall survival using data from two meta-analyses in early colon cancer (Sargent 2005) and advanced colon cancer (MAGIC 2004). In both cases, a very strong association was found at the individual level, clearly showing that the surrogate can explain more than 76% and 84% of our total uncertainty about the overall survival for advanced and early colon cancer, respectively.

At the trial level, even though large point estimates of the R^2_{trial} were obtained, the associated confidence intervals were relatively wide, hampering interpretation.

A number of additional issues require attention. First, when there are more than two arms in the clinical trials under consideration, one has the choice between calculating the validation measures using all arms simultaneously. Indeed, the information-theoretic developments carry through when Z represents a nominal covariate or, equivalently, a set of dummies, rather than a sole binary variable. Alternatively, the

measures can be calculated for every pair of arms deemed of interest. Second, when several trials are included into a meta-analysis, it is implicitly assumed that the arms are properly ordered. Such a situation arises, for example, when in all trials the control arms, on the one hand, and the active arms, on the other hand, are similar. Otherwise, application of the methodology can become quite cumbersome, or even arbitrary. Third, even though the measures provide a quantification of surrogacy, there remains the important question as to how large is large. It is tough to provide hard guidance and, arguably, decisions will have to be taken based on a number of quantitative and qualitative arguments combined. Finally, note that, by parsimoniously using information, the information-theoretic approaches may lead to tighter confidence intervals than in the hierarchical-model framework. This is an advantage, in addition to increased generality and flexibility.

Acknowledgement

We gratefully acknowledge support from Belgian IUAP/PAI network "Statistical Techniques and Modelling for Complex Substantive Questions with Complex Data".

6 References

Albert, A., Ioannidis, J.P.A., Reichelderfer, P., Conway, B., Coombs, R.W., Crane, L., Demasi, R., Dixon, D.O., Flandre, P., Hughes, M.D., Kalish, L.A., Larntz, K., Lin, D., Marschner, I.C., Munoz, A., Murray, J., Neaton, J., Pettinelli, C., Rida, W., Taylor, J.M.G., Welles, & S.L. (1998). Statistical issues for HIV surrogate endpoints: point/counterpoint. *Statistics in Medicine* **17**, 2435–2462.

Alonso A., and Molenberghs G. (2006). Surrogate Marker Evaluation from an

Information Theory Perspective. *Biometrics*, **00**, 000–000.

Alonso, A. Molenberghs, G., Burzykowski, T., Renard, D., Geys, H., Shkedy, Z., Tibaldi, F., Abrahantes, J., and Buyse, M. (2004). Prentice's approach and the meta analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics* **60**, 724–728.

Alonso, A., Geys, H., and Molenberghs, G. (2006) A unifying approach for surrogate marker validation based on Prentice's criteria. *Statistics in Medicine* **25**, 205–221.

Burzykowski, T., Molenberghs, G., and Buyse, M. (2005). The Evaluation of Surrogate Endpoints. New York: Springer.

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

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

Daniels, M.J. and Hughes, M.D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* **16**, 1515–1527.

Fisher, R. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* **22**, 700-725.

Freedman, L., Graubard, B., and Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* **11**, 167–178.

Gail, M., Pfeiffer, R., Van Houwelingen, H. & Carroll, R. (2000). On meta-analytic assessment of surrogate outcomes. *Biostatistics*, **1**, 231–246.

Hartley, R. (1928). Transmission of information. *Bell System Technical Journal* **7**, 535–563.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.

Meta-Analysis Group In Cancer (2004). Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: an updated meta-analysis. *Journal of Clinical Oncology* **22**, 3766–3775.

Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell System Technical Journal* **3**, 324–346.

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

Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002). Validation of surrogate endpoints in randomized trials with discrete outcomes. *Biometrical Journal* **44**, 1–15.

Renard D., Geys H., Molenberghs G., Burzykowski T., Buyse M., Vangeneugden T., Bijnens L. (2003). Validation of longitudinally measured surrogate marker for a time-to-event endpoint. *Journal of Applied Statistics* **30**, 235-247

Sargent DJ, Wieand S, Haller DG, Gray R, Benedetti J, Buyse M, Labianca R, Seitz JF, Callaghan CJO, Francini G, Grothey A, O'Connell M, Catalano PJ, Blanke CD, Kerr D, Green E, Wolmark N, Andre T, Goldberg RM, De Gramont A. (2005). Disease-free survival (DFS) vs. overall survival (OS) as a primary endpoint for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* **34**, 8664–8670,

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423 and 623–656.