Evaluation of Surrogate Endpoints Using Information-Theoretic Measure of Association Based on Havrda and Charvat Entropy

María del Carmen Pardo[1,†], **Qian Zhao**[2,†], **Hua Jin**[3], **Ying Lu**[4,*]

[1]Department of Statistics and O.R., Complutense University of Madrid, 28040 Madrid, Spain;

[2]Department of Epidemiology and Health Statistics, School of Public Health, Guangzhou Medical University, Guangzhou 510260, China;

[3]Department of Probability and Statistics, School of Mathematics, South China Normal University, Guangzhou 510631, China;

[4]Department of Biomedical Data Science, Stanford University, San Francisco, CA 94305-5464, USA

Abstract

Surrogate endpoints have been used to assess the efficacy of a treatment and can potentially reduce the duration and/or number of required patients for clinical trials. Using information theory, Alonso et al. (2007) proposed a unified framework based on Shannon entropy, a new definition of surrogacy that departed from the hypothesis testing framework. In this paper, a new family of surrogacy measures under Havrda and Charvat (H-C) entropy is derived which contains Alonso's definition as a particular case. Furthermore, we extend our approach to a new model based on the information-theoretic measure of association for a longitudinally collected continuous surrogate endpoint for a binary clinical endpoint of a clinical trial using H-C entropy. The new model is illustrated through the analysis of data from a completed clinical trial. It demonstrates advantages of H-C entropy-based surrogacy measures in the evaluation of scheduling longitudinal biomarker visits for a phase 2 randomized controlled clinical trial for treatment of multiple sclerosis.

Keywords

surrogate endpoint; information theory; Havrda and Charvat entropy; mutual information; clinical trial design

---

*Correspondence: ylu1@stanford.edu.
†The two authors are co-first authors.

1. Introduction

Surrogate endpoints which can be observed earlier, easier, possibly repeated, or are cost-saving, have been used to replace clinical endpoints in clinical trials. For example, total tumor response rate and progression–free survival have been used in phase II and phase III cancer clinical trials as surrogate endpoints for overall survival, which often requires a longer trial duration to achieve adequate statistical power. The United States Food and Drug Administration (USFDA) has accepted the use of surrogate endpoints in regulatory reviews of new drug applications [1]. Most cancer drug approvals (55 of 83 (66%)) between 2009 and 2014 by the USFDA have used at least one surrogate endpoint [2].

A motivative example is to use biomarkers in phase II cancer trials. Non-randomized single arm or randomized parallel clinical trials are used to evaluate signal of efficacy for a new drug. A binary response status, such as the total response based on the RECIST criterion [3], or a continuous response in change of tumor sizes [4] are common primary endpoints. For molecular-targeted drugs or immune oncology therapies, various serum, tissue, or imaging biomarkers are being developed to assess if the targeted pathways have been activated. These biomarkers are usually continuous, can be measured repeatedly, and their changes should proceed to a clinical response. However, the activation of a targeted pathway doesn't necessarily imply the response to the treatment. Various questions have been raised about the utility of such biomarkers in phase II trials [5,6].

Surrogate endpoints provide the convenience to speed up the clinical trial [7], but may not represent the actual outcomes well regarding the benefit of therapy [8]. For instance, bevacizumab was approved in metastatic breast cancer based on the surrogate outcome and was later withdrawn for failing to confirm a survival benefit [5]. How to evaluate the therapy benefit between surrogates, denoted as S and the true clinical endpoint outcome, denoted as T, remains a scientific challenge.

There are a lot of successful statistical methods and measures to assess surrogate endpoints. One method to validate surrogate endpoints is to evaluate their correlation with clinically meaningful endpoints through meta-analyses [9]. Only 11 of 89 (12%) studies had found high correlation (r 0.85), and nine (10%) showed a moderate-only correlation (r > 0.7 to r < 0.85) between surrogates and endpoints [7], suggesting that the strength of surrogates in clinical practice is often unknown or weak.

The landmark paper by Prentice [8] proposed operational criteria for the identification of valid surrogate endpoints. A sufficient condition for an endpoint S, as a valid surrogate of a primary clinical endpoint T, in the evaluation of a treatment, denoted as Z, is that the random vector (Z, S, T) forms a Markov chain $Z \rightarrow S \rightarrow T$, i.e., conditioning on S, Z and T are independent from each other. This condition led to parametric and non-parametric approaches to quantify the proportion of the treatment effect on T that is explained by the treatment effect on the S [9–13]. Other proposed quantities to assess the utility of a surrogate endpoint include dissociative effects, associative effects, average causal necessity, average causal sufficiency, causal effect predictiveness surface, and principle surrogate, etc. [9,14–22].

Buyse and Molenberghs [23] suggested two quantities to validate a surrogate endpoint: the relative effect which related the treatment effect on the primary outcome to that on the surrogate at the population level, and the adjusted association, which quantified the association between the primary outcome and surrogate marker after adjusting for the treatment at the individual level. These methods assumed that information regarding the surrogate and true endpoints was available from a single-trial surrogate evaluation method. They focused only on the validity, whereas the general association of the last one was related to the efficiency of a surrogate marker. Using an information-theoretic approach, Alonso and Molenberghs [24] and Pryseley et al. [25] redefined surrogacy in terms of the information content that S provides with respect to T. Using notations from Pryseley et al [25], let S be a continuous surrogate random variable and T be the continuous targeted clinical endpoint of interest. We use $f(\cdot)$ to denote the density function. The Shannon entropy functions for T and the conditional variable $T|S$ are denoted as $h(T)$ and $h(T|S)$, respectively, where $h(T) = E[-logf(T)]$ and $h(T|S) = E[-logf(T|S)]$. The corresponding entropy power functions are $EP_h(T) = e^{\frac{2}{n}h(T)}/(2\pi e)$ and $EP_h(T|S) = e^{\frac{2}{n}h(T|S)}/(2\pi e)$. An information theoretic measure of association (ITMA) is defined by Alonso and colleagues as the proportion of uncertainty reduction measured by the entropy power function for $T|S$ in reference to T:

$$R_h^2 = \frac{EP_h(T) - EP_h(T|S)}{EP_h(T)} = 1 - e^{-2I(T,S)} \qquad (1)$$

Here $I(T, S) = E[-logf(T)] - E[-logf(T|S)]$ is the mutual information. If S is a good surrogate for T, uncertainty about the effect on T is reduced by knowing the knowledge of the effect on S. There are some useful properties as described by Alonso and Molenberghs [24]:

1. $0 \leq R_h^2 \leq 1$

2. $R_h^2 = 0$, if and only if (T, S) are independent

3. R_h^2 is symmetric in (T, S)

4. R_h^2 is invariant under bijective transformations of T and S

5.   When $R_h^2 \rightarrow 1$ for continuous models, there is usually a deterministic relationship in the distribution of (T, S), that is, often $T = \varphi(S)$.

6.   When T is a discrete random variable, $R_h^2 \leq 1 - e^{-2h(T)}$. So they propose to use $R_{h,max}^2 = R_h^2/(1 - e^{-2h(T)})$ as a modified ITMA.

A good surrogacy should have a high R_h^2. The beauty of the informatic-theoretical framework is that it moves away from hypothesis testing and provides a quantitative measure of surrogacy. Thus, if there are two surrogate endpoints, S_1 and S_2, we can compare their utility as surrogacy endpoints based on the value of $R_h^{2'}$s.

Multiple authors have provided examples of this approach and demonstrated applications for situations when S and T are both binary, continuous, longitudinal, and time-to-event random variables as well as ordinal outcomes [24–28].

In this paper, we present two new results on the topic of surrogate endpoints based on information-theoretic measure of association (ITMA). First, we extend the ITMA construction based on Shannon entropy to a construction based on Havrda and Charvat (H-C) entropy [29]. The extension is motivated by the general existence of the H-C entropy. Explicit expressions as well as the properties of ITMA in different situations based on H-C entropy are presented. Second, we extend the H-C ITMA model to a longitudinally collected continuous surrogate endpoint for a binary clinical endpoint of a clinical trial. Then, the benefit of S is evaluated with the ITMA [24,25]. The current work focuses on a single trial surrogacy and its extension to a meta-analytic framework will need further development.

The paper is organized in the following structure. In Section 2, we give an example of when the Shannon entropy cannot be defined, thus the surrogacy by Alonso and Molenberghs under the information theoretical framework will not work [24]. We then prove the existence of H-C entropy under general conditions. Therefore, a family of surrogacy measures based on ITMA of H-C entropy is defined. An explicit formula is obtained for the following situations: binary-binary, continuous–continuous and binary-continuous. In Section 3, we extend a longitudinal linear random effects model for the longitudinally collected surrogate marker and a probit regression model for a binary primary endpoint in clinical trials. An application using H-C entropy in selecting times to collect surrogate measures is presented using data from a completed clinical trial. Finally, Section 4 presents discussions and a conclusion. R-Programs that generated re-sults for tables in Section 3 are presented in the Appendix A.

2. Extension of ITMA Surrogacy from Shannon Entropy to Havrda-Charvat Entropy

Why should we consider the extension? While Shannon's entropy is adequate in most applications, there are cases when a Shannon's entropy function doesn't exist, and thus the ITMA cannot be properly calculated. We give an example in the following.

Example 1.

Let X be the random variable with heavy tails. Its density function is

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & |x| \leq c_1 \\[2mm] \dfrac{c_2}{|x|(\log|x|)^2}, & |x| > c_1 \end{cases} \tag{2}$$

where $c_1 \approx 1.44$ and $c_2 \approx 0.027$ are chosen so that it is a continuous function with a heavy tail. The Shannon entropy $h(X)$ for X is infinity.

One way to make ITMA work is to use the Havrda-Charvat entropy [29], a generalization of entropy function that contains the Shannon entropy as its special case. Mathematically, Havrda-Charvat entropy is defined as follows:

$$HC_\alpha(X) = \int \varphi_\alpha(f(x))f(x)dx.$$ (3)

where

$$\varphi_\alpha(x) = \begin{cases} (x^{\alpha-1} - 1)/(1 - \alpha) & \alpha \neq 1 \\ -log(x) & \alpha = 1 \end{cases}$$ (4)

It is easy to see that $HC_1(X) = h(X)$.

Proposition 1.

Under a mild regular condition that the density function for X is bounded, $HC_\alpha(X)$ can always exist for an $\alpha > 1$.

Proof.

We only need to prove that for a bounded density function $f(x)$ by a constant $K > 0$, $\int (f(x))^\alpha dx < +\infty$.

Let $W = \{x: f(x) > 1\}$, then it follows from the fact that

$$1 = \int f(x)dx = \int_W f(x)dx + \int_{W^C} f(x)dx \geq m(W) + \int_{W^C} f(x)dx$$

that $m(W) \leq 1$ and $\int_{W^C} f(x)dx \leq 1$, where $m(W)$ is the probability of W. Therefore,

$$\int (f(x))^\alpha dx = \int_W (f(x))^\alpha dx + \int_{W^C} (f(x))^\alpha dx \leq K^{\alpha-1} m(W) \int_{W^C} f(x)dx < +\infty$$

Because Shannon's entropy is a special case of H-C entropy and H-C entropy always exists with proper choice of α, ITMA of H-C entropy should be a more flexible way as a surrogacy measure for more distribution families. Different from Shannon's entropy, H-C entropy satisfies a non-additive property such that $HC_\alpha(T, S) = HC_\alpha(T) + HC_\alpha(S) + (1 - \alpha)HC_\alpha(T)HC_\alpha(S)$, when T and S are independent. In general, the non-additive measures of entropy find justifications in many biological and chemical phenomena [30]. While H-C entropy has been used in quantum physics [31] and medical imaging research [32], it has not yet been used to describe the endpoint surrogacy for clinical trials.

To extend H-C entropy to measure the endpoint surrogacy for trials, we define the ITMA under H-C entropy power as the following:

$$R_\alpha^2 = 1 - e^{-2I_\alpha(T, S)} \tag{5}$$

where $I_\alpha(T, S)$ is the mutual information between T and S under H-C entropy [32]. Specifically, for $\alpha \neq 1$,

$$I_\alpha(T, S) = HC_\alpha(T) + HC_\alpha(S) + (1 - \alpha)HC_\alpha(T)HC_\alpha(S) - HC_\alpha(T, S) \tag{6}$$

When $\alpha = 1$, $I_\alpha(T, S) = I(T, S)$ and $R_\alpha^2 = R_h^2$ in Equation (1).

It is important to notice that: $\dfrac{EP_\alpha(T) - EP_\alpha(T \mid S)}{EP_\alpha(T)} \neq 1 - e^{-2I_\alpha(T, S)}$, for $\alpha \neq 1$ where

$EP_\alpha(X) = e^{\frac{2}{n}HC_\alpha(X)}/(2\pi e)$.

Some basic properties are available here.

1. When T and S are independent, $I_\alpha(T, S) = 0$. Thus, $R_\alpha^2 = 0$.

2. When T and S are deterministic, the value of $I_\alpha(T, S)$ will depend on $\alpha > 1$ or $\alpha < 1$ as seem in the following propositions.

Proposition 2.

Let T and S be two continuously normally distributed random variables such as

the joint distribution of $(T, S)' \sim N\left(\begin{bmatrix} \mu_T \\ \mu_S \end{bmatrix}, \begin{bmatrix} \sigma_T^2 & \rho\sigma_T\sigma_S \\ \rho\sigma_T\sigma_S & \sigma_S^2 \end{bmatrix} \right)$, the conditional distribution of

$(T \mid S) \sim N\left(\mu_T + \dfrac{\rho\sigma_T}{\sigma_S}(S - \mu_S), \sigma_T^2(1 - \rho^2) \right)$ and $T \sim N\left(\mu_T, \sigma_T^2 \right)$, where "$'$" means vector transpose.

Then, we have the following results:

2.1 The mutual information for H-C entropy depends not only the correlation between T and S, but also their standard deviations for $\alpha \neq 1$.

$$I_\alpha(T, S) = \frac{(2\pi\sigma_T\sigma_S)^{1 - \alpha}}{\alpha(1 - \alpha)}\left[1 - \left(\sqrt{1 - \rho^2} \right)^{1 - \alpha} \right] \quad \alpha \neq 1$$
$$I_\alpha(T, S) = I(T, S) = -\frac{1}{2}\log\left(1 - \rho^2 \right) for \, \alpha = 1 \tag{7}$$

2.2 When $\alpha \neq 1$, $\rho \rightarrow \pm 1$, $I_\alpha \rightarrow \infty$. $\rho \rightarrow \pm 0$, $I_\alpha \rightarrow 0$, I_α is an increasing function of $|\rho|$

2.3 When $\alpha < 1$, $\rho \rightarrow \pm 1$, $I_\alpha \rightarrow 1$. $\rho \rightarrow \pm 0$, $I_\alpha \rightarrow 0$, I_α is an increasing function of $|\rho|$.

Therefore, for $\alpha < 1$, maximum $R_\alpha^2 = 1 - e^{-2}$. We can normalize R_α^2 by dividing its

maximum value to make normalized $\widetilde{R}_\alpha^2$ in between 0 and 1

$$\widetilde{R}^2_{\alpha,\,max} = \frac{R^2_\alpha}{1 - e^{-2}}, \text{ for } \alpha < 1 \tag{8}$$

Proof.

$$\begin{aligned}
HC_\alpha(T, S) &= \int\int \frac{1}{(1-\alpha)}\left(f^{\alpha-1}_{T,S}(t, s) - 1\right)f_{T,S}(t, s)\,dt\,ds \\
&= \frac{1}{(1-\alpha)}\left[\int\int f^\alpha_{T,S}(t, s)\,dt\,ds - 1\right] \\
&= \frac{1}{(1-\alpha)}\left[\frac{2\pi\sigma_T\sigma_S\sqrt{1-\rho^2}}{\alpha\left(2\pi\sigma_T\sigma_S\sqrt{1-\rho^2}\right)^\alpha} - 1\right] \\
&= \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T\sigma_S\sqrt{1-\rho^2}\right)^{1-\alpha}}{\alpha} - 1\right]
\end{aligned}$$

Similarly,

$$HC_\alpha(T) = \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right]$$

$$HC_\alpha(S) = \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right]$$

And

$$\begin{aligned}
(1-\alpha)HC_\alpha(T)HC_\alpha(S) &= \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right]\left[\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right] \\
&= \frac{1}{(1-\alpha)}\frac{(2\pi\sigma_T\sigma_S)^{1-\alpha}}{\alpha} - \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} + \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\right] + \frac{1}{(1-\alpha)}
\end{aligned}$$

As such

$$I_\alpha(T,S) = \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right] + \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right] + \frac{1}{(1-\alpha)}\frac{(2\pi\sigma_T\sigma_S)^{1-\alpha}}{\alpha}$$

$$- \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} + \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\right] + \frac{1}{(1-\alpha)} - \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_T\sigma_S\sqrt{1-\rho^2}\right)^{1-\alpha}}{\alpha} - 1\right]$$

$$= \frac{(2\pi\sigma_T\sigma_S)^{1-\alpha}}{\alpha(1-\alpha)}\left[1 - \left(\sqrt{1-\rho^2}\right)^{1-\alpha}\right]$$

Finally, results in 2.2 and 2.3 can be concluded from the expression of $I_\alpha(T,S)$. $\square$

Proposition 3.

Let T and S be two binary outcome variables with 1 for a success and 0 for a failure such as the joint distribution of $(T,S)' \sim Multinomial(p_{0,0}, p_{1,0}, p_{0,1}, p_{1,1})$. We have the following results:

$$I_\alpha(T,S) = = \frac{1}{(1-\alpha)}\sum_{t=0}^1\sum_{s=0}^1\left\{\left[p_{t,+}\right]^\alpha\left[p_{+,s}\right]^\alpha - \left[p_{t,s}\right]^\alpha\right\} \tag{9}$$

3.1 When $p_{t,s} = p_{t,+}\,p_{+,s'}$ T and S are independent, $I_\alpha(T,S) = 0$.

3.2 Let $\rho = \frac{p_{1,1} - p_{1,+}p_{+,1}}{\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}}$ be the correlation between T and S. If $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1} > [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}$, $I_\alpha(T,S)$ is an increasing function of ρ for $\alpha > 1$. For $\alpha < 1$, $I_\alpha(T,S)$ is an increasing function of ρ if $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1} > [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}$.

3.3 For $\alpha > 1$, $I_\alpha(T,S) \quad \min(HC_\alpha(T), HC_\alpha(S))$.

3.4 For $\alpha < 1$, $I_\alpha(T,S) \quad \max(HC_\alpha(T), HC_\alpha(S))$.

3.5 For a given marginal distribution of T and S, there is a maximum value of mutual information as

$$I_\alpha(T,S) \le HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) - \min(HC_\alpha(T,S))$$
$$= I_{\alpha,max}. \tag{10}$$

Thus, we can normalize ITMA as $\widetilde{R}_{\alpha,max}(T,S) = \frac{R_\alpha(T,S)}{1 - e^{-2I_{\alpha,max}}}$.

Proof.

Because

$$HC_\alpha(T,S) = \frac{1}{(1-\alpha)}\sum_{t=0}^1\sum_{s=0}^1 p_{t,s}\left\{\left[p_{t,s}\right]^{\alpha-1} - 1\right\} = \frac{1}{(1-\alpha)}\left\{\sum_{t=0}^1\sum_{s=0}^1\left[p_{t,s}\right]^\alpha - 1\right\}$$

$$HC_\alpha(T) = \frac{1}{(1-\alpha)} \sum_{t=0}^{1} p_{t,+} \left\{ [p_{t,+}]^{\alpha-1} - 1 \right\} = \frac{1}{(1-\alpha)} \left\{ \sum_{t=0}^{1} [p_{t,+}]^{\alpha} - 1 \right\}$$

$$HC_\alpha(S) = \frac{1}{(1-\alpha)} \sum_{s=0}^{1} p_{+,s} \left\{ [p_{+,s}]^{\alpha-1} - 1 \right\} = \frac{1}{(1-\alpha)} \left\{ \sum_{s=0}^{1} [p_{+,s}]^{\alpha} - 1 \right\}$$

And

$$(1-\alpha)HC_\alpha(T)HC_\alpha(S) = \frac{1}{(1-\alpha)} \left\{ \sum_{t=0}^{1} [p_{t,+}]^{\alpha} - 1 \right\} \left\{ \sum_{s=0}^{1} [p_{+,s}]^{\alpha} - 1 \right\}$$

Thus,

$$HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) = \frac{1}{(1-\alpha)} \left[\left\{ \sum_{t=0}^{1} [p_{t,+}]^{\alpha} \right\} \left\{ \sum_{s=0}^{1} [p_{+,s}]^{\alpha} \right\} - 1 \right]$$

Therefore,

$$
\begin{aligned}
I_\alpha(T,S) &= \frac{1}{(1-\alpha)} \left(\left[\left\{ \sum_{t=0}^{1} [p_{t,+}]^{\alpha} \right\} \left\{ \sum_{s=0}^{1} [p_{+,s}]^{\alpha} \right\} - 1 \right] - \frac{1}{(1-\alpha)} \left\{ \sum_{t=0}^{1} \sum_{s=0}^{1} [p_{t,s}]^{\alpha} - 1 \right\} \right) \\
&= \frac{1}{(1-\alpha)} \left[\left\{ \sum_{t=0}^{1} [p_{t,+}]^{\alpha} \right\} \left\{ \sum_{s=0}^{1} [p_{+,s}]^{\alpha} \right\} - \sum_{t=0}^{1} \sum_{s=0}^{1} [p_{t,s}]^{\alpha} \right] \\
&= \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \sum_{s=0}^{1} \left\{ [p_{t,+}]^{\alpha} [p_{+,s}]^{\alpha} - [p_{t,s}]^{\alpha} \right\}
\end{aligned}
$$

which is the expression given in Equation (9).

Result 3.1 is the direct derivation from Equation (9) as $p_{t,s} = p_{t,+}p_{+,s}$.

Result 3.2 can be derived through the following relationship:

$$p_{1,1} = p_{1,+}p_{+,1} + \rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}$$

$$p_{1,0} = p_{1,+}p_{+,0} - \rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}$$

$$p_{0,1} = p_{0,+}p_{+,1} - \rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}$$

$$p_{0,0} = p_{0,+} + p_{+,0} + \rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}$$

where

$$\frac{max(-p_{1,+}p_{+,1}, -p_{0,+}p_{+,0}, p_{1,+}p_{+,0} - 1, p_{0,+}p_{+,1} - 1)}{\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}} \leq \rho$$
$$\leq \frac{min(1 - p_{1,+}p_{+,1}, 1 - p_{0,+}p_{+,0}, p_{1,+}p_{+,0}, p_{0,+}p_{+,1})}{\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}} \tag{11}$$

Then,

$$I_\alpha(T, S) = \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \sum_{s=0}^{1} \left\{ [p_{t,+}]^\alpha [p_{+,s}]^\alpha - [p_{t,+}p_{+,s} + (-1)^{t+s}\rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}}]^\alpha \right\}$$

Taking the derivative of $I_\alpha(T, S)$ on ρ, we have

$$\frac{dI_\alpha(T, S)}{d\rho} = \frac{\alpha}{(1-\alpha)} \sum_{t=0}^{1} \sum_{s=0}^{1} \Big\{$$
$$-(-1)^{t+s}\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}} \Big[p_{t,+}^{(z)}p_{+,s}^{(z)} + (-1)^{t+s}\rho\sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}} \Big]^{\alpha-1} \Big\}$$
$$= \frac{\alpha}{(1-\alpha)} \sum_{t=0}^{1} \sum_{s=0}^{1} \sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}} \Big\{ (-1)^{t+s+1}[p_{t,s}]^{\alpha-1} \Big\}$$
$$= \frac{\alpha}{(1-\alpha)} \sqrt{p_{1,+}p_{0,+}p_{+,1}p_{+,0}} \Big\{ -[p_{0,0}]^{\alpha-1} - [p_{1,1}]^{\alpha-1} + [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1} \Big\}$$

Under condition that $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1} > [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}, \frac{dI_\alpha(T, S)}{d\rho} > 0$ for $\alpha > 1$.

Similarly, under condition that $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1} < [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}, \frac{dI_\alpha(T, S)}{d\rho} > 0$ for $\alpha < 1$.

For 3.3 and 3.4

when $\alpha > 1$,

$$HC_\alpha(T) - I_\alpha(T, S) = \frac{1}{(\alpha-1)} \sum_{t=0}^{1} \left\{ 1 - [p_{t,+}]^\alpha - \sum_{s=0}^{1} \left\{ [p_{t,s}]^\alpha - [p_{t,+}]^\alpha [p_{+,s}]^\alpha \right\} \right\}$$
$$= \frac{1}{(\alpha-1)} \sum_{t=0}^{1} \left\{ 1 - \sum_{s=0}^{1} [p_{t,s}]^\alpha - [p_{t,+}]^\alpha \left\{ 1 - \sum_{s=0}^{1} [p_{+,s}]^\alpha \right\} \right\}$$
$$\geq \frac{1}{(\alpha-1)} \sum_{t=0}^{1} \left\{ 1 - \sum_{s=0}^{1} [p_{+,s}]^\alpha - [p_{t,+}]^\alpha \left\{ 1 - \sum_{s=0}^{1} [p_{+,s}]^\alpha \right\} \right\}$$
$$= \frac{1}{(\alpha-1)} \left\{ 1 - \sum_{s=0}^{1} [p_{+,s}]^\alpha \right\} \sum_{t=0}^{1} \left\{ 1 - [p_{t,+}]^\alpha \right\} \geq 0$$

However, when $\alpha < 1$,

$$HC_\alpha(T) - I_\alpha(T, S) = \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \left\{ [p_{t,+}]^\alpha - 1 - \sum_{s=0}^{1} \left\{ [p_{t,+}]^\alpha [p_{+,s}]^\alpha - [p_{t,s}]^\alpha \right\} \right\}$$

$$= \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \left\{ [p_{t,+}]^\alpha - 1 - [p_{t,+}]^\alpha \sum_{s=0}^{1} [p_{+,s}]^\alpha + \sum_{s=0}^{1} [p_{t,s}]^\alpha \right\}$$

$$\leq \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \left\{ [p_{t,+}]^\alpha - 1 - [p_{t,+}]^\alpha \sum_{s=0}^{1} [p_{+,s}]^\alpha + \sum_{s=0}^{1} [p_{+,s}]^\alpha \right\}$$

$$= \frac{1}{(1-\alpha)} \sum_{t=0}^{1} \left([p_{t,+}]^\alpha - 1 \right) \sum_{s=0}^{1} \left(1 - [p_{+,s}]^\alpha \right) \leq 0.$$

Because of a symmetric relationship between T and S, we proved results 3.4 and 3.5.

For 3.5, for fixed marginal probability, $I_\alpha(T, S)$ depends only on ρ in $HC_\alpha(T, S)$. Like 3.2,

$$\frac{dHC_\alpha(T, S)}{d\rho} = \frac{\alpha}{(1-\alpha)} \sqrt{p_{1,+} p_{0,+} p_{+,1} p_{+,0}} \left\{ -[p_{0,0}]^{\alpha-1} - [p_{1,1}]^{\alpha-1} + [p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1} \right\}$$

When $\frac{dHC_\alpha(T, S)}{d\rho} > 0$, taking the lower boundary of ρ in inequality (11) will derive the min value of $HC_\alpha(T, S)$.

When $\frac{dHC_\alpha(T, S)}{d\rho} < 0$, taking the upper boundary of ρ in inequality (11) will derive the min value of $HC_\alpha(T, S)$. □

Remark 1.

When the concordant pairs are more likely than the discordant pairs for the two binary endpoints, for $\alpha > 1$, $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1}$ is more likely to be greater than $[p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}$. However, when $\alpha < 1$, $[p_{0,0}]^{\alpha-1} + [p_{1,1}]^{\alpha-1}$ is more likely to be less than $[p_{1,0}]^{\alpha-1} + [p_{0,1}]^{\alpha-1}$. Thus, when two binary endpoints have more chance to be concordant, the mutual information will be an increasing function of correlation coefficient of ρ as shown in Proposition 3 *Result 3.2.*

Now we define a model for a binary *T* and a continuous normally distributed surrogate variable *S*.

Proposition 4.

Let T be a binary outcome variable and S continuous normally distributed surrogate variable, where $T \sim B(p_0, p_1)$ and $S \sim N(\mu_S, \sigma_S^2)$. We assume that there is a latent variable U such that $T = 1 \Leftrightarrow U \geq 0$, i.e, a Probit model with $U \sim N(\mu_T, 1)$ and $\mu_T = \Phi^{-1}(p_1)$. Assume a correlation coefficient between U and S is ρ, the conditional binary endpoint

T|S follows a Bernoulli distribution with $p_{1|s} = p_{T=1|S=s} = \Phi\left(\dfrac{\Phi^{-1}(p_1) + \rho\dfrac{s-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)$ *and*

$p_{0|s} = p_{T=0|S=s} = 1 - p_{T=1|S=s} = \Phi\left(\dfrac{\Phi^{-1}(p_0) - \rho\dfrac{s-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right).$ *We have the following results:*

4.1 *The mutual information for H-C entropy is*

$$I_\alpha(T,S) = \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^{1}[p_t]^\alpha - \int_{-\infty}^{\infty}\left[\sum_{t=0}^{1}[p_{t|s}]^\alpha\right]\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)ds\right\}$$

$$= \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{(1-\alpha)\sqrt{\alpha}}\sum_{t=0}^{1}\left[(p_t)^\alpha - \mathfrak{I}^{(\alpha)}\left(\frac{\Phi^{-1}(p_t)}{\sqrt{1-\rho^2}}, \frac{(-1)^{1-t}\dfrac{\rho}{\sqrt{\alpha}}}{\sqrt{1-\rho^2}}\right)\right]$$

where $\mathfrak{I}^{(\alpha)}(a,b) = \int_{-\infty}^{\infty}\Phi^\alpha(a+by)\phi(y)dy.$

4.2 *When* $\rho = 0$, $I_\alpha(T,S) = 0$.

4.3 *When* $\rho \to \pm 1$,

$$I_\alpha(T,S) \to \frac{1}{(1-\alpha)}\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\left\{\sum_{t=0}^{1}[p_t]^\alpha - \Phi\left(\sqrt{\alpha}\Phi^{-1}(p_1)\right) - \Phi\left(\sqrt{\alpha}\Phi^{-1}(p_0)\right)\right\}$$

4.4 *For* $\alpha \to 1$, $I_\alpha(T,S) \to \sum_{t=0}^{1}\left[\int_{-\infty}^{\infty}p_{t|s}log(p_{t|s})\phi(s)ds - p_t log(p_t)\right]$

Proof.

The joint distribution function for $(T, S) = (t, s)$ is:

$$f(t, s) =$$

$$\left\{ \int_0^\infty \frac{1}{2\pi\sigma_S\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}\left[u-\mu_T \quad \frac{S-\mu_S}{\sigma_S}\right]\begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}^{-1}\begin{bmatrix}u-\mu_T \\ \frac{S-\mu_S}{\sigma_S}\end{bmatrix}\right)du\right\}^t$$

$$\times\left\{ \int_{-\infty}^0 \frac{1}{2\pi\sigma_S\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}\left[u-\mu_T \quad \frac{S-\mu_S}{\sigma_S}\right]\begin{bmatrix}1 & \rho\sigma_S \\ \rho\sigma_S & \sigma_S^2\end{bmatrix}^{-1}\begin{bmatrix}u-\mu_T \\ \frac{S-\mu_S}{\sigma_S}\end{bmatrix}\right)du\right\}^{1-t}$$

$$=\left\{ \int_0^\infty \frac{1}{2\pi\sigma_S\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left(\left[u-\mu_T-\rho\frac{S-\mu_S}{\sigma_S}\right]^2+\left(1-\rho^2\right)\left[\frac{S-\mu_S}{\sigma_S}\right]^2\right)\right)du\right\}^t$$

$$\times\left\{ \int_{-\infty}^0 \frac{1}{2\pi\sigma_S\sqrt{1-\rho^2}} exp\left(-\frac{1}{2(1-\rho^2)}\left(\left[u-\mu_T-\rho\frac{S-\mu_S}{\sigma_S}\right]^2+\left(1-\rho^2\right)\left[\frac{S-\mu_S}{\sigma_S}\right]^2\right)\right)du\right\}^{1-t}$$

$$=\left[1-\Phi\left(\frac{-\mu_T-\rho\frac{S-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^t\left[\Phi\left(\frac{-\mu_T-\rho\frac{S-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^{1-t}\frac{\phi\left(\frac{S-\mu_S}{\sigma_S}\right)}{\sigma_S}$$

$$=\left[1-\Phi\left(\frac{-\Phi^{-1}(p_1)-\rho\frac{S-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^t\left[\Phi\left(\frac{-\Phi^{-1}(p_1)-\rho\frac{S-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^{1-t}\frac{\phi\left(\frac{S-\mu_S}{\sigma_S}\right)}{\sigma_S}$$

$$=[p_{1|s}]^t[p_{0|s}]^{1-t}\frac{\phi\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S}$$

Therefore,

$$HC_\alpha(T) = \frac{1}{(1-\alpha)}\left\{\sum_{t=0}^1 [p_t]^\alpha - 1\right\}$$

$$HC_\alpha(S) = \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right]$$

$$HC_\alpha(T, S) = \frac{1}{1-\alpha}\left\{\sigma_S^{1-\alpha}\int_{-\infty}^\infty [p_{1|s}^\alpha + p_{0|s}^\alpha]\phi^\alpha\left(\frac{S-\mu_S}{\sigma_S}\right)d\frac{S}{\sigma_S} - 1\right\}$$

$$HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) = \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^1 [p_t]^\alpha - 1\right\}$$

$$I_\alpha(T, S)$$
$$= HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) - HC_\alpha(T, S)$$

$$= \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^1 [p_t]^\alpha - \int_{-\infty}^\infty\left[\sum_{t=0}^1 p_{t|s}\alpha\right]\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S^\alpha}ds\right\}$$

Therefore,

$$HC_\alpha(T) = \frac{1}{(1-\alpha)}\left\{\sum_{t=0}^{1}[p_t]^\alpha - 1\right\}$$

$$HC_\alpha(S) = \frac{1}{(1-\alpha)}\left[\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} - 1\right]$$

$$HC_\alpha(T, S) = \frac{1}{1-\alpha}\left\{\sigma_S^{1-\alpha}\int_{-\infty}^{\infty}[p_{1|s}{}^\alpha + p_{0|s}{}^\alpha]\phi^\alpha\left(\frac{S-\mu_S}{\sigma_S}\right)d\frac{S}{\sigma_S} - 1\right\}$$

$$HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) = \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^{1}[p_t]^\alpha - 1\right\}$$

$$
\begin{aligned}
I_\alpha&(T, S)\\
&= HC_\alpha(T) + HC_\alpha(S) + (1-\alpha)HC_\alpha(T)HC_\alpha(S) - HC_\alpha(T, S)\\
&= \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^{1}[p_t]^\alpha - \int_{-\infty}^{\infty}\left[\sum_{t=0}^{1}p_{t|s}{}^\alpha\right]\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds\right\}
\end{aligned}
$$

Using $\mathfrak{F}^{(\alpha)}(a, b) = \int_{-\infty}^{\infty}\Phi^\alpha(a + by)\phi(y)dy$, we can derive an alternative formulation

$$\int_{-\infty}^{\infty}\left[\sum_{t=0}^{1}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\rho\frac{s-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^\alpha\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds$$

$$= \int_{-\infty}^{\infty}\left[\sum_{t=0}^{1}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\rho\frac{S-\mu_S}{\sigma_S}}{\sqrt{1-\rho^2}}\right)\right]^\alpha\frac{e^{-\frac{\alpha}{2\sigma_S^2}(s-\mu_S)^2}}{\left(2\pi\sigma_S^2\right)^{\frac{\alpha}{2}}}ds$$

$$= \frac{\left(2\pi\sigma_S^2/\alpha\right)^{\frac{1}{2}}}{\left(2\pi\sigma_S^2\right)^{\frac{\alpha}{2}}}\int_{-\infty}^{\infty}\left[\sum_{t=0}^{1}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)\right]^\alpha\frac{e^{-\frac{1}{2}y^2}}{(2\pi)^{\frac{1}{2}}}dy$$

$$= \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^{1}\int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)^\alpha\phi(y)dy$$

$I_\alpha(T, S)$ can be simplified as

$$I_\alpha(T, S) = \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{(1-\alpha)\sqrt{\alpha}}\sum_{t=0}^{1}\left[(p_t)^\alpha - \mathfrak{F}^{(\alpha)}\left(\frac{\Phi^{-1}(p_t)}{\sqrt{1-\rho^2}}, \frac{(-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}}{\sqrt{1-\rho^2}}\right)\right]$$

Thus, we complete the proof for 4.1.

For 4.2, $\rho = 0$,

$$I_\alpha(T,S) = \frac{1}{(1-\alpha)}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\sum_{t=0}^{1}[p_t]^\alpha - \int_{-\infty}^{\infty}[p_1{}^\alpha + p_0{}^\alpha]\frac{\phi^\alpha\left(\frac{S-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds\right\} = 0$$

For 4.3, as $\rho \to 1$,

$$\int_{-\infty}^{\infty}[p_{1|s}{}^\alpha + p_{0|s}{}^\alpha]\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds \to \int_{-\Phi^{-1}(p_1)}^{\infty}\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds + \int_{-\infty}^{\Phi^{-1}(p_0)}\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds$$

$$= \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\left[\Phi\left(\sqrt{\alpha}\Phi^{-1}(p_1)\right) + \Phi\left(\sqrt{\alpha}\Phi^{-1}(p_0)\right)\right]$$

Similarly, $\rho \to -1$,

$$\int_{-\infty}^{\infty}[p_{1|s}{}^\alpha + p_{0|s}{}^\alpha]\frac{\phi^\alpha\left(\frac{s-\mu_S}{\sigma_S}\right)}{\sigma_S{}^\alpha}ds \to \frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\left[\Phi\left(\sqrt{\alpha}\Phi^{-1}(p_1)\right) + \Phi\left(\sqrt{\alpha}\Phi^{-1}(p_0)\right)\right].$$

So, $\rho \to \pm 1$, $I_\alpha(T,S) \to \frac{1}{(1-\alpha)}\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}}\left\{\sum_{t=0}^{1}\left[(p_t)^\alpha - \Phi\left(\sqrt{\alpha}\Phi^{-1}(p_t)\right)\right]\right\}.$

For $\alpha = 1$, H-C entropy is similar to Shannon's entropy. Thus, by taking limit of α to 1, we can derive Shannon's mutual information for the Probit model in 4.4.

$$\lim_{\alpha \to 1} I_\alpha(T,S) = \lim_{\alpha \to 1}\left\{\frac{\left(2\pi\sigma_S^2\right)^{\frac{1-\alpha}{2}}}{(1-\alpha)\sqrt{\alpha}}\sum_{t=0}^{1}\left[(p_t)^\alpha - \int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)^\alpha\phi(y)dy\right]\right\}$$

$$= \lim_{\alpha \to 1}\left\{\frac{1}{(1-\alpha)}\sum_{t=0}^{1}\left[(p_t)^\alpha - \int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)^\alpha\phi(y)dy\right]\right\}$$

$$= \lim_{\alpha \to 1}\left\{\frac{1}{-1}\sum_{t=0}^{1}\left[(p_t)^\alpha log(p_t) - \int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)^\alpha log\left(\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\frac{\rho}{\sqrt{\alpha}}y}{\sqrt{1-\rho^2}}\right)\right)\phi(y\right.\right.$$

$$\left.\left.)dy\right]\right\}$$

$$= \sum_{t=0}^{1}\left[\int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\rho y}{\sqrt{1-\rho^2}}\right)log\left(\Phi\left(\frac{\Phi^{-1}(p_t) + (-1)^{1-t}\rho y}{\sqrt{1-\rho^2}}\right)\right)\phi(y)dy - p_t log(p_t)\right]$$

$$= \sum_{t=0}^{1}\left[\int_{-\infty}^{\infty}p_{t|y}log(p_{t|y})\phi(y)dy - p_t log(p_t)\right]$$

Remark 2.

Taking into account that $\mathfrak{F}^{(2)}(a, b) = \Phi\left(\dfrac{a}{\sqrt{1 + b^2}}\right) - 2T\left(\dfrac{a}{\sqrt{1 + b^2}}, \dfrac{1}{\sqrt{1 + 2b^2}}\right)$, *where*

$T(h, k) = \phi(h)\int_0^k \dfrac{\phi(hx)}{1 + x^2}dx$ *is the Owen's function* [33] *and the property* $T(h, k) = T(-h, k)$, *we can derive explicit formula for* $a=2$,

$$I_2(T, S) = -\frac{1}{2\sqrt{\pi}\sigma_S}\sum_{t=0}^{1}\left[(p_t)^2 - \left[\Phi\left(\frac{\Phi^{-1}(p_t)}{\sqrt{1 - \rho^2/2}}\right) - 2T\left(\frac{\Phi^{-1}(p_t)}{\sqrt{1 - \rho^2/2}}, \sqrt{(1 - \rho^2)}\right)\right]\right]$$

Since $\Phi^{-1}(p_t) = \Phi^{-1}(1 - p_{1-t}) = -\Phi^{-1}(p_{1-t})$ *and* $\Phi\left(\dfrac{\Phi^{-1}(p_1)}{\sqrt{1 - \rho^2/2}}\right) + \Phi\left(\dfrac{\Phi^{-1}(p_0)}{\sqrt{1 - \rho^2/2}}\right) = 1$ *we can*

simplify the expression as

$$I_2(T, S) = \frac{1}{2\sqrt{\pi}\sigma_S}\left\{1 - 4T\left(\frac{\Phi^{-1}(p_0)}{\sqrt{1 - \rho^2/2}}, \sqrt{(1 - \rho^2)}\right) - (p_1)^2 - (p_0)^2\right\}$$

3. Surrogacy of a Longitudinal Biomarker for a Binary Clinical Endpoint

3.1. Model for Longitudinal Continuous Surrogate Biomarkers in Phase II Trials

In many phase II trials, clinical endpoints of interest (T) are often a proportion of a binary endpoint or mean of a continuous variable. For example, in oncology phase II trials, a common clinical endpoint is total response rate. The surrogate biomarkers, on the other hand, are usually lab tests either from serum or urine or imaging modalities that can be measured repeatedly during the study. In this section, we focus on a binary one-time clinical endpoint T and a continuous repeated surrogate variable S.

In the remainder of the paper, we will use t_j to denote the time of jth measurement, since baseline in a longitudinal trial. For simplicity, consider the difference model from baseline $t_0 = 0$.

Let the general model as:

$$S_{i,j}|Z_i, T_i = \mu_{S_i} + \alpha_1 Z_i + \alpha_2 t_j + \alpha_3 Z_i t_j + \beta_j T_i + \epsilon_{S, ij}, \text{ for } j = 1, ..., K.$$

Thus, $S_i|Z_i, T_i = (S_{i,0}|Z_i, T_i, S_{i,1}|Z_i, T_i, ..., S_{i,K}|Z_i, T_i)' \sim MVN(\mu_{S_i}, \Sigma_{SS})$, where

$$\mu_{S_i} = (\mu_S + \alpha_1 Z_i + (\alpha_2 + \alpha_3 Z_i)t_1 + \beta_1 T_i, ..., \mu_S + \alpha_1 Z_i + (\alpha_2 + \alpha_3 Z_i)t_K + \beta_K T_i)'$$

Using a bivariate probit model [34] for the joint distribution of $(Z_i, T_i)'$ we can derive a probit model for the conditional joint distribution for $(Z_i, T_i)'|S$:

$$\left[\begin{matrix} \Phi^{-1}[P(T_i = 1)] \\ \Phi^{-1}[P(Z_i = 1)] \end{matrix}\middle| S_i\right] \sim MVN_2\left(\begin{bmatrix} \mu_{0,1} \; \gamma_1' \\ \mu_{0,2} \; \gamma_2' \end{bmatrix}\begin{bmatrix} 1 \\ S_i \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

where $\mu_{0,k}$ and γ_k are the intercept and regression coefficient vector for the probit regression for T given longitudinal S ($k = 1$) and for Z given longitudinal S ($k = 2$), respectively, and ρ is the correlation coefficient of two underlying latent normal variables for T and Z.

Because a linear combination of multivariate normal variables is still a normal random variable, we can use the Proposition 4 to calculate ITMA under H-C entropy power to evaluate surrogacy of the longitudinal biomarker in each arm conditioning on Z. We can also average over the treatment arms to get the mean trial level ITMA under H-C entropy, denoted as

$$I_\alpha(T, S|Z) = E[I_\alpha(T, S|Z)] = I_\alpha(T, S|Z = 1)P(Z = 1) + I_\alpha(T, S|Z = 0)P(Z = 0)$$

Furthermore, we can use the mutual information $I_\alpha(T, Z|S)$ to verify Prentice's criteria as suggested by [24]: i.e, conditioning on surrogate S, the clinical endpoint T and treatment assignment Z are independent, thus a good surrogate should lead to $I_\alpha(T, Z|S) \approx 0$. Since

$$I_\alpha(T, Z|S) = E\langle HC_\alpha(T|S) + HC_\alpha(Z|S) + (1 - \alpha)HC_\alpha(T|S)HC_\alpha(Z|S) - HC_\alpha(T, Z|S)\rangle$$

where

$$HC_\alpha(T|S) = \frac{1}{1 - \alpha}\left\{[p(T = 1|S)]^\alpha + [p(T = 0|S)]^\alpha - 1\right\} \tag{12}$$

$$HC_\alpha(Z|S) = \frac{1}{1 - \alpha}\left\{[p(Z = 1|S)]^\alpha + [p(Z = 0|S)]^\alpha - 1\right\}$$

$$HC_\alpha(T, Z|S) = \frac{1}{1 - \alpha}\left\{[p(T = 1, Z = 1|S)]^\alpha + [p(T = 1, Z = 0|S)]^\alpha + [p(T = 0, Z = 1|S)]^\alpha + [p(T = 0, Z = 0|S)]^\alpha - 1\right\}$$

So for $\alpha \neq 1$,
$$I_\alpha(T, Z|S) = \frac{1}{1 - \alpha}E\left\langle\sum_{t=0}^1\sum_{z=0}^1\left\{[p(T = t|S)]^\alpha[p(Z = z|S)]^\alpha - [p(T = t, Z = z|S)]^\alpha\right\}\right\rangle$$ When $\alpha = 1$, $I_\alpha(T, Z|S) = E\langle-\sum_{t=0}^1 p(T = t|S)log[p(T = t|S)] - \sum_{z=0}^1 p(Z = z|S)log[p(Z = z|S)].$
$+ \sum_{t=0}^1\sum_{z=0}^1 p(T = t, Z = z|S)log[p(T = t, Z = z|S)]\rangle$

For real data, we can use bivariate probit model to estimate equations for $[p(T = t|S)]^\alpha$, $[p(Z = z|S)]^\alpha$, and $[p(T = t, Z = z|S)]^\alpha$, then use Equation (9) to perform numerical integration for the derivation of $I_\alpha(T, Z|S)$. One way to perform this analysis is to use R-package mvProbit

from CRAN-R (https://cran.r-project.org/web/packages/mvProbit/mvProbit.pdf, accessible on January 29, 2022).

3.2. A Data Example

"Safety, Tolerability and Activity Study of Ibudilast in Subjects with Progressive Multiple Sclerosis" (NCT01982942) is a US National Institute of Health (NIH) sponsored multicenter, randomized, double-blind, placebo-controlled, parallel-group phase II study from November 2013 to December 2017. The main study results have been published by Fox, et al. [35]. The trial data is publicly available upon request to NIH. We use this data for the numerical illustration for H-C ITMA.

More specifically, patients were enrolled with primary or secondary progressive multiple sclerosis of this phase II randomized trial of oral ibudilast (100 mg daily) or placebo for 96 weeks. The primary efficacy end point was the rate of brain atrophy, as measured by the brain parenchymal fraction (brain size relative to the volume of the outer surface contour of the brain). Major secondary end points included the change in the pyramidal tracts on diffusion tensor imaging and cortical atrophy, all measures of tissue damage in multiple sclerosis.

We requested and received data from the study team that contained 104 placebo patients and 99 treated patients, with longitudinal observations in brain parenchymal fraction (BPF) and thinning of the cortical gray matter (cortical thickness) measured by magnetic resonance imaging at week 0, 24, 48, 72, and 96. For illustration purposes, we altered the primary and secondary endpoints of the trial and created a binary clinical endpoint as the cortical thickness (CTH) greater than 3 mm as a clinical outcome for less cortical gray matter atrophy and used BPF as the continuous longitudinal marker. Table 1 provides a summary of the data used for this illustration.

From Table 1, we can see that 104 patients were randomized to the control arm and 99 patients to the treatment arm. The treatment significantly reduced cortical atrophy for 71% patients who maintained more than 3 mm cortical thickness (CTH) in the treatment arm in comparison to 48% in the placebo arm at 96 weeks post baseline. While the differences in BPFs between treatment arms had p-values above 0.38 in each follow-up MRI, the aggregated changes over time measured by the slopes of a mixed random effects regression model achieved highly statistical significance with a p-value of 0.0056.

The importance of evaluating the surrogacy of the longitudinal BPF measurements for the binary CTH endpoint in MS trials is to understand the strength of surrogacy and whether it can be used to shorten trial duration. More importantly for future trial design, we need to understand how often and when the longitudinal measurements should be performed.

Using formulas derived in Proposition 4, we derived the mean mutual information and ITMA of longitudinal BPF as a surrogate for the clinical endpoint of maintaining more than 3 mm cortical thickness at the end of 96 weeks. We explore three choices of $a = 0.5$, 1, and 2 to show the difference between H-C and Shannon entropies. The value of $a = 1$ has been considered because it corresponds to Shannon entropy. The other two alpha values have been

considered in other papers such as [32]. The columns of Table 2 are organized according to values of a. Each row in Table 2 represents a design to use BPF in the baseline (week 0) and different follow-up visits to construct a longitudinal surrogate endpoint. For example, the first row used the baseline and week 24 data while the last row used the data from baseline, weeks 48 and 72.

From Table 2, we can see that the longitudinal BPF measures at the baseline with at least one follow-up visit were all reasonable surrogates for the binary endpoint of CTH > 3.0 mm. H-C entropy with $a = 0.5$ was not sensitive enough to differentiate subtle differences in surrogacy utility of different designs to collect surrogate endpoints. When $a = 1$, H-C entropy is Shannon entropy and it was able to discriminate among different designs to the 3rd decimal place. H-C entropy for $a = 2$ was more sensitive and showed differences in all designs. As it demonstrated, using longitudinal BPF data could shorten trial duration to 72 weeks. For a trial ended at 72 weeks, additional BPF measures at week 24 and week 48 did not add any more valuable utility to surrogacy than a single measure in week 24. Overall, the p-values from the linear mixed random effects model reflected the directions of ITMAs, but not in completely concordance, perhaps, due to random variation in fitting the mixed random effects and the probit models.

Table 3 examines the longitudinal surrogacy based on Prentice's criteria. Here we want to determine if $I_a(T, Z|S)$ is close to 0. The results of Table 3 confirm the observations in Table 2 that the longitudinal BPF is a good surrogate variable for binary CTH > 3.0 mm at 96 weeks. Because Table 3 uses the same model as Table 2, the p-values for longitudinal models are omitted. Once again, $I_a(T, Z|S)$ decreases with a.

4. Conclusions

Alonso et al. [24] proposed to assess the validity of a surrogate endpoint in terms of uncertainty reduction. The main proposals for measures of uncertainty are found in information theory. These authors based their proposal in the well-known Shannon entropy. In the past there has been an extensive work on generalized entropies [30–32,36–39]. We focus on the Havrda-Charvat entropy, which reduces to the Shannon case if the parameter is set to one, to extend that surrogacy measure. Based on the generalized entropy, we consider a generalized mutual information as it has been proved in other contexts to have better performance of some members of this family [30–32]. In this paper, the theoretical development of these measures has been completed. The advantage of our proposal is that it contains a particular case of a useful measure to assess surrogacy and demonstrates the ability to easily explore other measures which may have performance advantages for specific questions. We have seen the advantage of using $a = 2$ instead of $a = 1$ in our example to evaluate scheduling of longitudinal visits.

Some additional issues are pending. On one hand, we are working to carry out a more extensive numerical study for assessing the performance of these measures in the endpoint surrogacy context. In our paper, we compared the performance of ITMA in a real trial with three choices of a (0.5, 1 and 2). They were chosen for illustration purposes. The optimal choice of a remains a research question. Additional research can consider other ITMA, such

as divergence measures [36], taking into account that the mutual information is equal to the Kullback divergence, or measures of unilateral dependency as that defined by Andonie et al. [37] based on the informational energy [39] or surrogacy for testing of variances [38].

Acknowledgments:

Data Availability Statement:

Data can be requested through the US National Institute of Neurological Disorders and Stroke and Fox, the Principal Investigator of NCT01982942 for access the de-identified data from the "Safety, Tolerability and Activity Study of Ibudilast in Subjects with Progressive Multiple Sclerosis."

Appendix A.: R-Program

R-Program for Table 1

```
### Row 1 ###
fisher.test(table(nihexample$CTh96YesNo,nihexample$trt.group))
### Row 3-5 ###
t.test(nihexample$bpf0~nihexample$trt.group)
t.test(nihexample$bpf96~nihexample$trt.group)
### Row 6 ###
library(lme4)
summary(lmer(bpf~trt.group+week+trt.group*week+(1|ID),data=nihexamplelong))
```

R-Program for Table 2

```
hcentr=function(pt,preds,pz,trt, alpha){
s2=var(preds)
if(alpha!=1){
mtrinf=pz/(1-alpha)*((2*pi*s2)^((1-alpha)/2)/sqrt(alpha)*((pt[2])^alpha+(1-
pt[2])^alpha)-mean((pnorm(preds[trt==1]))^alpha+(1-
pnorm(preds[trt==1]))^alpha))+(1-pz)/(1-alpha)*((2*pi*s2)^((1-alpha)/2)/
sqrt(alpha)*((pt[1])^alpha+(1-pt[1])^alpha)-
mean((pnorm(preds[trt==0]))^alpha+(1-pnorm(preds[trt==0]))^alpha))}
```

```
if(alpha==1){ mtrinf=pz*(mean(pnorm(preds[trt==1])*log(pnorm(preds[trt==1])))
+mean(1-pnorm(preds[trt==1])*log(1-pnorm(preds[trt==1])))-
pt[2]*log(pt[2])-(1-pt[2])*log(1-pt[2]))+(1-
pz)*(mean(pnorm(preds[trt==0])*log(pnorm(preds[trt==0])))+mean(1-
pnorm(preds[trt==0])*log(1-pnorm(preds[trt==0])))-pt[1]*log(pt[1])-(1-
pt[1])*log(1-pt[1]))}
itma=1-exp(-2*mtrinf)
c(mtrinf,itma)
}
### Row 1 ###
pt=table(nihexample$CTh96YesNo,nihexample$trt.group)[2,]/
table(nihexample$trt.group)
myprobit1=glm(CTh96YesNo~trt.group+bpf0+bpf24,family=binomial(link="probit"),
 data=nihexample)
row1=round(c(
hcentr(pt,myprobit1$linear.predictors,sum(nihexample$trt.group)/
length(nihexample$trt.group),nihexample$trt.group,0.5),
hcentr(pt,myprobit1$linear.predictors,sum(nihexample$trt.group)/
length(nihexample$trt.group),nihexample$trt.group,1),
hcentr(pt,myprobit1$linear.predictors,sum(nihexample$trt.group)/
length(nihexample$trt.group),nihexample$trt.group,2),
summary(lmer(bpf~trt.group+week+trt.group*week+(1|
ID),data=nihexamplelong[nihexamplelong$week==0 |nihexamplelong$week==24, ]))
$coefficients[4,5]),4)
### similar codes for other rows ####
```

R-Program for Table 3

```
library(mvtnorm)
library(mvProbit)
####(T, Z)####
table(table(nihexample$CTh96YesNo,nihexample$trt.group)
#####choose varibles: "bpf0", "bpf24" "bpf48" "bpf72" "bpf96" (1-5)
fullmodel1 =
mvProbit(cbind(CTh96YesNo,trt.group)~bpf0+bpf24+bpf48+bpf72+bpf96,data=nihexa
mple)
#####model
summary(fullmodel1)
sigma=symMatrix(c(1,fullmodel1$estimate[length(fullmodel1$estimate)],1))
#############################################
#mu1=fullmodel1$estimate[1]+fullmodel1$estimate[2]*nihexample$bpf0+fullmodel1
$estimate[3]*nihexample$bpf24
#mu2=fullmodel1$estimate[4]+fullmodel1$estimate[5]*nihexample$bpf0+fullmodel1
$estimate[6]*nihexample$bpf24
```

```
#mu1=fullmodel1$estimate[1]+fullmodel1$estimate[2]*nihexample$bpf0+fullmodel1
$estimate[3]*nihexample$bpf24+fullmodel1$estimate[4]*nihexample$bpf48
#mu2=fullmodel1$estimate[5]+fullmodel1$estimate[6]*nihexample$bpf0+fullmodel1
$estimate[7]*nihexample$bpf24+fullmodel1$estimate[8]*nihexample$bpf48
#mu1=fullmodel1$estimate[1]+fullmodel1$estimate[2]*nihexample$bpf0+fullmodel1
$estimate[3]*nihexample$bpf24+fullmodel1$estimate[4]*nihexample$bpf48+fullmod
el1$estimate[5]*nihexample$bpf72
#mu2=fullmodel1$estimate[6]+fullmodel1$estimate[7]*nihexample$bpf0+fullmodel1
$estimate[8]*nihexample$bpf24+fullmodel1$estimate[9]*nihexample$bpf48+fullmod
el1$estimate[10]*nihexample$bpf72
mu1=fullmodel1$estimate[1]+fullmodel1$estimate[2]*nihexample$bpf0+fullmodel1$
estimate[3]*nihexample$bpf24+fullmodel1$estimate[4]*nihexample$bpf48+fullmode
l1$estimate[5]*nihexample$bpf72+fullmodel1$estimate[6]*nihexample$bpf96
mu2=fullmodel1$estimate[7]+fullmodel1$estimate[8]*nihexample$bpf0+fullmodel1$
estimate[9]*nihexample$bpf24+fullmodel1$estimate[10]*nihexample$bpf48+fullmod
el1$estimate[11]*nihexample$bpf72+fullmodel1$estimate[12]*nihexample$bpf96
fullmodel1$estimate
sigma
#################### T|S Z|S############
bs=10000
set.seed(873465)
BT=sample(c(1:length(CTh96YesNo)), size = bs, replace = TRUE) ### bootstrap
ID###
BT_U1=matrix(0,bs,2) #### 1 D normal probability
BT_U2=matrix(0,bs,4) ####2 D normal probability
for (i in 1:bs)
{ c=BT[i]
BT_U1[i,1]=1-pnorm(0, mu1[c], 1) ####P(T=1|S)
BT_U1[i,2]=1-pnorm( 0,mu2[c], 1)####P(Z=1|S)
BT_U2[i,1]=pmvnorm(lower=c(0,0),up-
per=Inf,mean=c(mu1[c],mu2[c]),sigma)####P(T=1,Z=1|S)
BT_U2[i,2]=pmvnorm(lower=c(0,-Inf),up-
per=c(Inf,0),mean=c(mu1[c],mu2[c]),sigma)####P(T=1,Z=0|S)
BT_U2[i,3]=pmvnorm(lower=c(-
Inf,0),upper=c(0,Inf),mean=c(mu1[c],mu2[c]),sigma)####P(T=0,Z=1|S)
BT_U2[i,4]=pmvnorm(lower=-
Inf,upper=c(0,0),mean=c(mu1[c],mu2[c]),sigma)####P(T=0,Z=0|S)}
###############################
alfa=0.5
p1=mean((BT_U1[,1]*BT_U1[,2])^alfa+((1-BT_U1[,1])*BT_U1[,2])^alfa+
(BT_U1[,1]*(1-BT_U1[,2]))^alfa+((1-BT_U1[,1])*(1-BT_U1[,2]))^alfa)
p2=mean(BT_U2[,1]^alfa+BT_U2[,2]^alfa+BT_U2[,3]^alfa+BT_U2[,4]^alfa)
I_alfa=1/(1-alfa)*(p1-p2)
IM_alfa=1-exp(-2*I_alfa)
```

```
I_alfa
IM_alfa
###################################
alfa=2
p1=mean((BT_U1[,1]*BT_U1[,2])^alfa+((1-BT_U1[,1])*BT_U1[,2])^alfa+
(BT_U1[,1]*(1-BT_U1[,2]))^alfa+((1-BT_U1[,1])*(1-BT_U1[,2]))^alfa)
p2=mean(BT_U2[,1]^alfa+BT_U2[,2]^alfa+BT_U2[,3]^alfa+BT_U2[,4]^alfa)
I_alfa=1/(1-alfa)*(p1-p2)
IM_alfa=1-exp(-2*I_alfa)
I_alfa
IM_alfa
######## alfa=1####################
p1=-mean(BT_U1[,1]*log(BT_U1[,1])+(1-BT_U1[,1])*log(1-BT_U1[,1]))
p2=-mean(BT_U1[,2]*log(BT_U1[,2])+(1-BT_U1[,2])*log(1-BT_U1[,2]))
p3=mean(BT_U2[,1]*log(BT_U2[,1])+BT_U2[,2]*log(BT_U2[,2])
+BT_U2[,3]*log(BT_U2[,3]) +BT_U2[,4]*log(BT_U2[,4]))
I=p1+p2+p3
IM=1-exp(-2*I)
```

References

1. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. U.S. Department of Health and Human Services. Available online: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics fda.gov (accessed on January 29, 2022).

2. Kim C; Prasad V Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: An analysis of 5 years of, U.S.; food and drug administration approvals. JAMA Intern. Med 2015, 175, 1992–1994. 10.1001/jamainternmed.2015.5868. [PubMed: 26502403]

3. Schwartz LH; Litière S; de Vries E; Ford R; Gwyther S; Mandrekar S; Shankar L; Bogaerts J; Chen A; Dancey J; et al. RECIST 1.1-update and clarification: From the RECIST committee. Eur. J. Cancer 2016, 62, 132–137. 10.1016/j.ejca.2016.03.081. [PubMed: 27189322]

4. Karrison TG; Maitland ML; Stadler WM; Ratain MJ Design of phase II cancer trials using a continuous endpoint of change in tumor size: Application to a study of sorafenib and erlotinib in non-small-cell lung cancer. J. Natl. Cancer Inst 2007, 99, 1455–1461; Erratum in J. Natl. Cancer Inst. 2007, 99, 1819. [PubMed: 17895472]

5. Burzykowski T; Coart E; Saad ED; Shi Q; Sommeijer DW; Bokemeyer C; Díaz-Rubio E; Douillard JY; Falcone A; Fuchs CS; et al. Evaluation of continuous tumor-size–based end points as surrogates for overall survival in randomized clinical trials in metastatic colorectal cancer. JAMA Netw. Open 2019, 2, e1911750. 10.1001/jamanetworkopen.2019.11750. [PubMed: 31539075]

6. Lu Y Statistical considerations for quantitative imaging measures in clinical trials. In Biopharmaceutical Applied Statistics Symposium: Volume 3 Pharmaceutical Applications; Peace KE, Chen D-G, Menon S, Eds.; ICSA Book Series in Statistics; Springer: Singapore, 2018; pp. 219–240.

7. Chen EY; Joshi SK; Tran A; Prasad V Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. JAMA Intern. Med 2019, 179, 642–647. 10.1001/jamainternmed.2018.8351. [PubMed: 30933235]

8. Kok PS; Yoon WH; Lord S; Marschner I; Friedlander M; Lee CK Tumor response end points as surrogates for overall survival in immune checkpoint inhibitor trials: A systematic review and meta-analysis. JCO Precis. Oncol 2021, 5, 1151–1159. 10.1200/PO.21.00108.

9. Shameer K; Zhang Y; Jackson D; Rhodes K; Neelufer IKA; Nampally S; Prokop A; Hutchison E; Ye J; Malkov VA; et al. Correlation between early endpoints and overall survival in non-small-cell

lung cancer: A trial-level meta-analysis. Front. Oncol 2021, 11, 672916. 10.3389/fonc.2021.672916. [PubMed: 34381708]

10. Haslam A; Hey SP; Gill J; Prasad V A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. Eur. J. Cancer 2019, 106, 196–211. 10.1016/j.ejca.2018.11.012. [PubMed: 30528804]

11. Prentice RL Surrogate endpoints in clinical trials: Definitions and operational criteria. Stat. Med 1989, 8, 431–440. [PubMed: 2727467]

12. Freedman LS; Graubard BI; Schatzkin A Statistical validation of intermediate endpoints for chronic diseases. Stat. Med 1992, 11, 167–178. [PubMed: 1579756]

13. Wang Y; Taylor JM A measure of the proportion of treatment expect explained by a surrogate marker. Biometrics 2002, 58, 803–812. [PubMed: 12495134]

14. Taylor JM; Wang Y; Thiffebaut R Counterfactual links to the proportion of treatment effect explained by a surrogate marker. Biometrics 2005, 61, 1102–1111. [PubMed: 16401284]

15. Parast L; Tian L; Cai T Landmark estimation of survival and treatment effect in a randomized clinical trial. J. Am. Stat. Assoc 2014, 109, 384–394. 10.1080/01621459.2013.842488. [PubMed: 24659838]

16. Parast L; McDermott MM; Tian L Robust estimation of the proportion of treatment effect explained by surrogate marker information. Stat. Med 2016, 35, 1637–1653. 10.1002/sim.6820. [PubMed: 26631934]

17. Frangakis CE; Rubin DB Principal stratification in causal inference. Biometrics 2002, 58, 21–29. [PubMed: 11890317]

18. Conlon AS; Taylor JM; Elliott MR Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. Biostatistics 2014, 15, 266–283. [PubMed: 24285772]

19. Huang Y; Gilbert PB Comparing biomarkers as principal surrogate endpoints. Biometrics 2011, 67, 1442–1451. [PubMed: 21517791]

20. Gabriel EE; Gilbert PB Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. Biostatistics 2014, 15, 251–265. [PubMed: 24337534]

21. Gabriel EE; Sachs MC; Gilbert PB Comparing and combining biomarkers as principle surrogates for time-to-event clinical endpoints. Stat. Med 2015, 34, 381–395. [PubMed: 25352131]

22. Gilbert PB; Hudgens MG Evaluating candidate principal surrogate endpoints. Biometrics 2008, 64, 1146–1154. [PubMed: 18363776]

23. Buyse M; Molenberghs G Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 1998, 54, 1014–1029. [PubMed: 9840970]

24. Alonso A; Molenberghs G Surrogate marker evaluation from an information theoretic perspective. Biometrics 2007, 63, 180–186. [PubMed: 17447943]

25. Pryseley A; Tilahun A; Alonso A; Molenberghs G Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. Clin. Trials 2007, 4, 587–597. [PubMed: 18042568]

26. Alonso A; Molenberghs G Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. Stat. Methods Med. Res 2008, 17, 497–504. [PubMed: 18285443]

27. Alonso A; Bigirumurame T; Burzykowski T; Buyse M; Molenberghs G; Muchene L; Perualila NJ; Shkedy Z; Van der Elst W Applied surrogate endpoint evaluation methods with SAS and R. Chapman Hall 2017. 10.1201/9781315372662

28. Ensor H; Weir CJ Evaluation of surrogacy in the multi-trial setting based on information theory: An extension to ordinal outcomes. J. Biopharm. Stat 2020, 30, 364–376. 10.1080/10543406.2019.1696357. [PubMed: 31887069]

29. Havrda J; Charvát F Quantification method of classification processes. Concept of structural α-entropy. Kybernetika 1967, 3, 30–35.

30. Tsallis C Possible generalization of BoltzmannGibbs statistics. J. Stat. Phys 1988, 52, 479–487.

31. Amigó JM; Balogh SG; Hernández S A brief review of generalized entropies. Entropy 2018, 20, 813.

32. Wachowiak MP; Smolíková R; Tourassi GD; Elmaghraby AS Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration. In Medical Imaging 2003: Image Processing; International Society for Optics and Photonics: Bellingham, WA, USA, 2003; Volume 5032, pp. 1090–1100.

33. Owen D A table of normal integrals. Commun. Stat. Simul. Comput 1980, 9, 389–419.

34. Chib S; Greenger E Analysis of multivariate probit models, Biometrika 1998, 85, 347–361.

35. Fox RJ; Coffey CS; Conwit R; Cudkowicz ME; Gleason T; Goodman A; Klawiter EC; Matsuda K; McGovern M; Naismith RT; et al. NN102/SPRINT-MS trial investigators. Phase 2 trial of ibudilast in progressive multiple sclerosis. N. Engl. J. Med 2018, 379, 846–855. 10.1056/NEJMoa1803583. [PubMed: 30157388]

36. Biswas A; Pardo MC; Guha A Auto-association measures for stationary time series of categorical data. TEST 2004, 23, 487–514.

37. Andonie R; Petrescu F Interacting systems and informational energy. Found. Control Eng 1986, 11, 53–59.

38. Pardo JA; Pardo MC; Vicente ML; Esteban MD A statistical information theory approach to compare the homogeneity of several variances. Comput. Stat. Data Anal 1997, 24, 411–416.

39. Menéndez ML; Pardo JA; Pardo MC Estimators based on sample quantiles using (h,φ)-entropy measures. Appl. Math. Lett 1998, 11, 99–104.

Table 1.

Summary Statistics for The Real Data Example.

Variable	Control (N = 104)	Treatment (N = 99)	*p*-Value [*]
CTH > 3 mm: N (%)	50 (48%)	70 (71%)	0.0016
BPF: Mean (SD)			
Week 0	0.8023 (0.0301)	0.8040 (0.0281)	0.6823
Week 24	0.8012 (0.0301)	0.8039 (0.0277)	0.5001
Week 48	0.8009 (0.0311)	0.8036 (0.0282)	0.5115
Week 72	0.8001 (0.0303)	0.8032 (0.0283)	0.4433
Week 96	0.7989 (0.0306)	0.8026 (0.0293)	0.3813
Change/24 weeks [**]	−0.0008 (0.0001)	−0.0004 (0.0001)	0.0056

[*] *p*-value for CTH > 3 mm was calculated using Fisher's exact test; *p*-values for mean differences at follow-up visits were calculated using a *t*-test. P-value for changes in 24 weeks (slopes) was calculated by the mixed random effects model.

[**] change per 24 weeks was estimated using a mixed random effects linear regression model using the R-lmer package.

Table 2.

H-C Mutual Information and ITMA by Different Longitudinal Designs.

BPF Data Used	$a = 0.5$		$a = 1$		$a = 2$		p-Value [*]
	$I_a(T, S\|Z)$	ITMA	$I_a(T, S\|Z)$	ITMA	$I_a(T, S\|Z)$	ITMA	
0, 24	4.6042	0.9999	2.6300	0.9948	0.6063	0.7026	0.0797
0, 24, 48	4.6117	0.9999	2.6307	0.9948	0.6066	0.7028	0.1025
0, 24, 48, 72	4.6209	0.9999	2.6352	0.9949	0.6071	0.7031	0.0390
0, 24, 48, 72, 96	4.6103	0.9999	2.6361	0.9949	0.6069	0.7029	0.0056
0, 48	4.4683	0.9999	2.6012	0.9945	0.5980	0.6976	0.1586
0, 72	4.4522	0.9999	2.5912	0.9944	0.5962	0.6965	0.0675
0, 24, 72	4.6223	0.9999	2.6348	0.9949	0.6072	0.7031	0.0485
0, 48, 72	4.4696	0.9999	2.6022	0.9945	0.5980	0.6976	0.0382

[*] p-value for treatment and visit interactions in a linear mixed random effects model using the R-lmer function.

Table 3.

Prentice Criteria for Surrogate Endpoint.

BPF Data Used	$a = 0.5$		$a = 1$		$a = 2$	
	$I_a(T, Z\|S)$	ITMA	$I_a(T, Z\|S)$	ITMA	$I_a(T, Z\|S)$	ITMA
0, 24	0.0390	0.0751	0.0271	0.0528	0.0108	0.0213
0, 24, 48	0.0388	0.0747	0.0270	0.0526	0.0108	0.0213
0, 24, 48, 72	0.0407	0.0782	0.0280	0.0545	0.0110	0.0218
0, 24, 48, 72, 96	0.0395	0.0760	0.0274	0.0533	0.0110	0.0218
0, 48	0.0428	0.0820	0.0297	0.0578	0.0117	0.0231
0, 72	0.0416	0.0798	0.0287	0.0558	0.0111	0.0219
0, 24, 72	0.0403	0.0775	0.0278	0.0541	0.0110	0.0217
0, 48, 72	0.0434	0.0832	0.0299	0.0580	0.0116	0.0229